



**UNIVERZA V LJUBLJANI**  
**FAKULTETA ZA RAČUNALNIŠTVO IN INFORMATIKO**

Avtor: mag. Karl Petrič

**Model adaptivne digitalne knjižnice na podlagi izdelave interesnih  
profilov uporabnikov**  
**DOKTORSKA DISERTACIJA**

Mentor: izr. prof. dr. Marjan Krisper

Somentor: red. prof. dr. Vladislav Rajkovič

**Ljubljana, 21.05.2008**

# **☐ Kazalo vsebine**

## **☐ 1 Uvod 10**

**☐ 1.1 Predstavitev problematike 10**

**☐ 1.2 Uporabniško usmerjen pristop k problematiki raziskovanja uporabnikov na digitalnih (spletnih) knjižnicah 11**

**☐ 1.3 Zgradba doktorskega dela 12**

**☐ 1.4 Cilj doktorskega dela 13**

**☐ 1.5 Opis problematike 14**

**☐ 1.6 Pričakovani prispevki k znanosti 14**

**☐ 1.7 Metodologija in metodološka orodja 16**

**☐ 1.8 Raziskovalne hipoteze in vprašanja 16**

**☐ 1.8.1 Raziskovalne hipoteze 16**

**☐ 1.8.2 Raziskovalna vprašanja 16**

## **☐ 2 Teoretična izhodišča 18**

**☐ 2.1 Webometrija (spletometrija) 18**

**☐ 2.1.1 Zakon moči (Power Law) 19**

**☐ 2.1.2 Zipfov zakon 20**

**☐ 2.1.2.1 Zakon o frekvenčni distribuciji korenov besed 20**

**☐ 2.2 Ontologija 22**

**☐ 2.2.1 Univerzalna decimalna klasifikacija (UDK) 23**

**☐ 2.2.2 Globalni pomen UDK področij (izpeljane osnovne logične značilnosti) 24**

**☐ 2.2.3 Pravila globalnega razvrščanja poizvedb po UDK 27**

**☐ 2.2.4 Nekaj primerov globalnega UDK razvrščanja na podlagi spletometričnih podatkov iz prakse 27**

**☐ 2.2.5 Tezaver 31**

**☐ 2.3 Odkrivanje zakonitosti v podatkih 32**

**☐ 2.3.1 Nekatere metode in tehnike podatkovnega rudarjenja ali odkrivanja zakonitostih v podatkih 33**

**☐ 2.3.2 Klasifikacija 33**

**☐ 2.3.3 Asociacijska pravila 34**

**☐ 2.3.4 Razvrščanje v skupine 35**

<b>2.3.5 Vizualizacijske tehnike 35</b>
2.3.5.1 Distribucija 35
2.3.5.2 Statistični atributi 35
2.3.5.3 Scatter Plot 36
2.3.5.4 Sievov diagram 36
2.3.5.5 Dendrogram 36
2.3.5.6 Distančna mapa 36
<b>2.3.6 Odkrivanje zakonitosti podatkov v besedilih 36</b>
2.3.6.1 Pomembnost ali moč določene ključne besede 37
<b>2.4 Digitalne knjižnice 38</b>
<b>2.5 Izdelava interesnih profilov uporabnikov na podlagi preučevanih dejavnosti uporabnikov na pilotni digitalni knjižnici 39</b>
2.5.1 Spletni iskalniki 39
<b>3 Analiza za pripravo modela 42</b>
<b>3.1 Zasnova modela 42</b>
<b>3.2 Načrtovanje podatkov – globalni pogled 43</b>
<b>3.3 Načrtovanje podatkov – osredotočen pogled 45</b>
3.3.1 Uporabljeni programska orodja 46
3.3.2 Nadzorna plošča spletnega notranjega iskalnika Pico Search 46
3.2.4 Orange Canvas in Antconc 48
3.3 Določitev virov podatkov in opis 50
3.4 Kratko poročilo o podatkih 50
3.6 Kratko poročilo o kakovosti pridobljenih podatkov 51
<b>3.7 Priprava podatkov 52</b>
3.7.1 Druga stopnja ali priprava besedilnega dela podatkov 57
3.8.2 Opis podatkov 58
3.8.3 Kratko poročilo o čiščenju podatkov 58
3.9 Tretja stopnja priprave podatkov 59
<b>4 Odkrivanje zakonitosti v podatkih v okviru modela 66</b>
4.3 Izhodišče 77
4.4 Distribucija in statistični atributi poizvedb 77
4.6 Scatterplot za število poizvedb po UDK področjih 81

<b>4.7 Prikaz podakovnih vzorcev s pomočjo Sievovega diagrama</b>	<b>86</b>
<b>4.8 Odkrivanje zanimivih vzorcev v podatkih z ozirom na poizvedbe s pomočjo dejavnika pomembnosti ali moči določene ključne besede</b>	<b>89</b>
<b>4.9 Izpeljana asociacijska pravila</b>	<b>96</b>
<b>4.9.5 Hierarhično grozdenje asociacijskih pravil in ključnih besed</b>	<b>106</b>
<b>4.9.7 Ovrednotenje podatkov (ugotovitev)/izpeljava novih znanj</b>	<b>113</b>
<b>4.9.9 Segmentacija spletnih obiskovalcev/poizvedovalcev</b>	<b>123</b>
<b>5 Kvalitativni model adaptivne digitalne knjižnice in ovrednotenje</b>	<b>136</b>
<b>5.1 Možnost uporabe izpeljanih novih znanj/spoznanj</b>	<b>136</b>
<b>5.3 Adaptivna digitalna knjižnica in multidimenzionalni rang tezaver (MRT)</b>	<b>141</b>
<b>5.4 Adaptivna digitalna knjižnica (ADK), multidimenzionalni rang tezaver (MRT) in ovrednotenje modela</b>	<b>142</b>
<b>6 Sklepno poglavje</b>	<b>151</b>
<b>6.1 Kakšni so bili izidi?</b>	<b>152</b>
<b>6.1.1 Zakon moči pred in po čiščenju podatkov</b>	<b>152</b>
<b>6.1.2 Distribucija in statistični atributi za poizvedbe</b>	<b>152</b>
<b>6.1.3 Primerjava frekvenc znotraj UDK področij z vsemi frekvencami</b>	<b>153</b>
<b>6.1.4 Razpon števila poizvedb znotraj UDK področij</b>	<b>153</b>
<b>6.1.5 Prikaz podakovnih vzorcev s pomočjo Sievovega diagrama</b>	<b>154</b>
<b>6.2 Odkrivanje zanimivih vzorcev v podatkih z ozirom na poizvedbe s pomočjo dejavnika pomembnosti ali moči določene ključne besede</b>	<b>154</b>
<b>6.2.1 Distribucija K, ranga, statistični atributi dejavnika K in stolpčni diagram UDK</b>	<b>154</b>
<b>6.2.2 Primerjava dejavnika K in r s pomočjo Scatterplota</b>	<b>154</b>
<b>6.2.3 Vizualizacija ključnih besed na podlagi dejavnika K</b>	<b>155</b>
<b>6.2.4 Izpeljana asociacijska pravila</b>	<b>156</b>
<b>6.2.5 Hierarhično grozdenje pomembnih ključnih besed</b>	<b>158</b>
<b>6.2.6 Najpomembnejša izpeljana nova znanja/spoznanja</b>	<b>159</b>
<b>7 Zaključek</b>	<b>165</b>
<b>7.1 Pogled v prihodnostni razvoj adaptivnih digitalnih knjižnic (ADK)</b>	<b>165</b>
<b>8 Priloga</b>	<b>166</b>
<b>8.1 Slovar izrazov</b>	<b>166</b>

<b>9 Viri in uporabljena programska orodja 169</b>
9.1 Priporočeni viri na podlagi posebnih izvedenih poizvedb s ključno besedo »MULTIDIMENSIONAL THESAURUS« na Web of Science, INSPEC in ERIC 175
9.2 Priporočeni viri s področja adaptivnih digitalnih knjižnic z zbirk podatkov Web of Science (WOS), ERIC, LISA, Compendex in INSPEC od leta 2000 do 2007 177
9.3 Uporabljena programska orodja 185
<b>10 Vsebina CD-ROM-a 185</b>
Izjava 186
Zahvala 187

<b>Kazalo tabel</b>
2.1.2.2 Preglednica 1: Zakon moči pri poizvedbah (2002 do 2005) 21
3.7.2 Preglednica 2: Rangi, frekvence in logaritmizirane frekvence znotraj posameznih UDK razredov 53
3.7.3 Preglednica 3: Rangirana UDK področja in logaritmizirane frekvence za leto 2003, 2004, 2005 in za vse skupaj 54
4.1 Preglednica 4: Zakon moči pred in po čiščenju podatkov 69
4.5 Preglednica 5: Primerjavava med frekvencami znotraj UDK področij in vsemi frekvencami 78
4.9.2 Preglednica 6: Asociacijska pravila besednih podatkov na podlagi posebej pripravljene .tra datoteke 99
4.9.6.1 Preglednica 7: Primerjava med dejavnikom najpomembnejših ključnih besed in števila obiskov spletnih obiskovalcev po UDK področjih 109
4.9.9.2 Preglednica 8: Matrika segmentacije spletnih obiskovalcev v skupine 130
5.7 Preglednica 9: Vrednotenje in kritična presoja modela z ozirom na uporabnika 148

## **▣ Kazalo slikovnih prikazov**

**▣ 2.1.2.3 Slikovni prikaz 1: Primerjava logaritmiranih frekvenc z rangi 21**

**▣ 2.2.5.1 Slikovni prikaz 2: Možna različica odnosov med pojmi v tezavru 31**

**▣ 3.1.1 Slikovni prikaz 3: Snovanje kvalitativnega modela modela adaptivne digitalne knjižnice z multidimenzionalnim rang tezavrom 42**

**▣ 3.2.1 Slikovni prikaz 4: Spletometrija poizvedb in postopki obdelave strukturiranih podatkov – načrtovanje podatkov 43**

**▣ 3.3.3 Slikovni prikaz 5: Zajem poizvedb, ki so jih izvedli spletni obiskovalci 45**

**▣ 3.3.5 Slikovni prikaz 6: Programsko orodje za OZVP Orange Canvas 48**

**▣ 3.3.6 Slikovni prikaz 7: Programsko orodje AntConc za OZVP v besedilih 49**

**▣ 3.7.1 Slikovni prikaz 8: Globalni pogled na pripravo podatkov za OZVP 52**

**▣ 3.7.4 Slikovni prikaz 9: Zakon moči in preizkusi 54**

**▣ 3.8.1 Slikovni prikaz 10: Čiščenje poizvedb s pomočjo programskega orodja AntConc 58**

**▣ 3.9.1 Slikovni prikaz 11: Priprava podatkov s pomočjo programskega orodja Orange Canvas 60**

**▣ 3.9.2 Slikovni prikaz 12: Priprava podatkov za ugotavljanje asociacijskih povezav in hierarhičnih grozdov 63**

**▣ 4.1.1 Slikovni prikaz 13: Logaritmirani in po rangih razvrščeni podatki 69**

**▣ 4.2 Slikovni prikaz 14: Okvirna miselna hierarhija interesnih področij spletnih obiskovalcev – poizvedovalcev na UDK leksikonu 73**

**▣ 4.4.1 Slikovni prikaz 15: Distribucija znotraj UDK področij in statistični atributi 77**

**▣ 4.5.1 Slikovni prikaz 16: Distribucija frekvenc znotraj UDK področij in UDK področij s frekvencami 78**

**▣ 4.5.2 Slikovni prikaz 17: Krivulje verjetnosti in zaupanja znotraj posameznih UDK področij – kakovostni vidik 80**

**▣ 4.6.1 Slikovni prikaz 18: Razpon števila poizvedb znotraj UDK področij 81**

**▣ 4.6.2 Slikovni prikaz 19: Porazdelitev razpona poizvedb z ozirom na skupno število poizvedb znotraj posameznih UDK področij in frekvenčna porazdelitev 84**

**▣ 4.7.1 Slikovni prikaz 20: Gostost poizvedb za posamezna UDK področja in skupaj 86**

**▣ 4.7.2 Slikovni prikaz 21: Zavzeta površina poizvedb po posameznih UDK področjih 87**

<input type="checkbox"/> 4.7.3 Slikovni prikaz 22: Model sobe oziroma poslopja 88
<input type="checkbox"/> 4.8.1 Slikovni prikaz 23: Distribucija K, ranga, statistični atributi dejavnika K in stolpčni diagram UDK 90
<input type="checkbox"/> 4.8.2 Slikovni prikaz 24: Primerjava dejavnika K in r s pomočjo Scatterplota 91
<input type="checkbox"/> 4.8.3 Slikovni prikaz 25: Vizualizacija ključnih besed na podlagi dejavnika K 94
<input type="checkbox"/> 4.9.1 Slikovni prikaz 26: Postopek izpeljave asociacijskih pravil na podlagi dejavnika K, r, N-ja, UDK in ključnih besed 97
<input type="checkbox"/> 4.9.3 Slikovni prikaz 27: Vzorec porazdelitve asociacijskih povezav 102
<input type="checkbox"/> 4.9.4 Slikovni prikaz 28: Distančne mape za dejavnika podpore in zaupanja ter ocenjena moč asociacijskih povezav in pravil 104
<input type="checkbox"/> 4.9.6 Slikovni prikaz 29: Dendrograma hierarhičnih grozdov za asociacijska pravila 107
<input type="checkbox"/> 4.9.6.2 Slikovni prikaz 30: Hierarhično in K – means grozdenje 109
<input type="checkbox"/> 4.9.7.1 Slikovni prikaz 31: Zakon moči poizvedb, primerjava med številom poizvedb in obiskov ter trendne črte 114
<input type="checkbox"/> 4.9.7.2 Slikovni prikaz 32: Porazdelitev poizvedb in trend zakona moči 116
<input type="checkbox"/> 4.9.8 Slikovni prikaz 33: Model sobe in pomembne ključne besede 118
<input type="checkbox"/> 4.9.8.1 Slikovni prikaz 34: Okrnjena semantična mreža hierarhije zanimanj 120
<input type="checkbox"/> 4.9.9.1 Slikovni prikaz 35: Pregled (izobraževalnih) zanimanj spletnih obiskovalcev/poizvedovalcev 123
<input type="checkbox"/> 4.9.9.3 Slikovni prikaz 36: Značilnosti miselnih svetov spletnih obiskovalcev/poizvedovalcev 132
<input type="checkbox"/> 5.2 Slikovni prikaz 37: Moduli adaptivne digitalne knjižnice (ADK) 139
<input type="checkbox"/> 5.3.1 Slikovni prikaz 38: Kvalitativen model adaptivne digitalne knjižnice 141
<input type="checkbox"/> 5.5 Slikovni prikaz 39: Zgradba multidimenzionalnega tezavra (MRT) in izsek prototipa tezavra 144
<input type="checkbox"/> 5.6 Slikovni prikaz 40: Odnosi med dimenzijami 146

## **Podatki o doktorskem delu**

**Ime in priimek:** Karl Petrič

**Naslov doktorskega dela:** Model adaptivne digitalne knjižnice na podlagi izdelave interesnih profilov uporabnikov

**Kraj:** Ljubljana

**Leto:** 2008

**Število strani:** 187 **Število slikovnih prikazov:** 40 **Število preglednic:** 9

**Število opomb:** 42 **Število navedenih virov:** 142 **Spremno gradivo:** 1 CD-ROM

**Mentor:** Izred. Prof. Dr. Marjan Krisper **Somentor:** Red. Prof. Dr. Vladislav Rajkovič

**UDK** 02:004.738.5

**Ključne besede:** webometrija (*spletometrija*), spletometrične analize, (*spletno*) odkrivanje zakonitosti v podatkih, digitalne knjižnice, poizvedbe, spletni obiskovalci, ontologije, organizacija informacij, profiliranje uporabnikov, spletno izobraževanje.

**Izveček:** V teoretičnem uvodu sem se ukvarjal z opredelitvijo spletometrije (*vključujoče metode zakona moči*) in nekaterih metod s področja odkrivanja zakonitosti v podatkih (*OZVP*). V nadaljevanju doktorskega dela sem se ukvarjal z analizo poizvedb, ki so jih izvedli spletni obiskovalci v obdobju od 01.01.2003 do 01.01.2006 na notranjem iskalniku pilotne različice digitalne knjižnice. S pomočjo metode »zakona moči« pri obdelavi podatkov sem ugotovil stalni informacijski/poizvedovalni vedenjski vzorec, ki se tudi v daljšem časovnem obdobju ni spreminjal. V nadaljevanju tega dela sem podatke še obdelal z nekaterimi metodami *OZVP* (*text mining, hierarhično grozdenje, asociacijska pravila ipd.*). Prav na podlagi opravljenih meritev in izdelanih analiz sem lahko razvil določene miselne modele spletnih obiskovalcev na splošnoizobraževalni ravni. Na koncu tega dela sem še predstavil lastni adaptivni model digitalne knjižnice.

**Title of the doctoral dissertation:** A model of an adaptive digital library on the basis of creating user interest profiles

**UDC** 02:004.738.5

**Keywords:** webometrics, webometric analyses, web usage mining, digital libraries, search retrievals, web users, user profiling, onthology, web education

**Abstract:** The introduction of the doctoral thesis deals with different definitions of webometrics and some methods of data mining. In the following chapters search retrievals of web users from 01/01/2003 to 01/01/2006 on the internal search engine of the pilot digital

library have been analysed. With the “Power Law” method of data processing a constant information retrieval pattern has been established, stable over a longer period of time. In the subsequent chapters the data have been processed with some other methods of data mining (*i.e. text mining, association rules etc.*). On the basis of the accomplished measurements and analyses a series of mind models of web users for general educational purposes have been developed. The doctoral thesis concludes with the construction of the author’s own adaptive model of a digital library.

# 1 Uvod

## 1.1 Predstavitev problematike

Samozadostni in prekomerni institucionalni značaj knjižnic se tudi v današnjem času rabe sodobne informacijske tehnologije ni tako bistveno spremenil kot bi lahko pričakovali. V zadnjem desetletju se sicer mnogo poroča o digitalnih knjižnicah, ki pa še vedno bolj kot ne nastopajo v vlogi izoliranega korelata pod taktirko bibliotekarske znanosti in stroke. Prav zaradi tega so digitalne knjižnice mnogokrat interpretirane kot zbirke podatkov ali celo kot seznam dokumentov z določenega interesnega področja. S pomočjo sodelovanja drugih znanosti in aplikativnih ved, h katerim naj še zlasti prištejem računalništvo in informatiko, so digitalne knjižnice ne zgolj sistematično urejene zbirke znanja, dostopne preko računalniškega komunikacijskega medija, ampak so to v končni stopnji informacijski sistemi za upravljanje z znanjem, ki zmorejo posredovati različne pomembne odločitve (*npr. poslovne odločitve, odločitve v zvezi z izobraževalnim procesom*) in naj bi povrh tega služile kot odlične raziskovalne platforme za raziskovanje vedenjskih vzorcev uporabnikov.

Raziskovanje vedenjskih vzorcev uporabnikov digitalnih knjižnic je eden od predpogojev, da je sploh možno vzpostaviti oziroma izdelati adaptivne digitalne knjižnice, ki se (*kot že samo ime pove*) prilagajajo potrebam različnih vrst uporabnikov (*notranji uporabniki, zunanji uporabniki itd.*). Adaptivne digitalne knjižnice (*v nadaljevanju: ADK*) morajo biti podprte s sodobno informacijsko tehnologijo morajo (*npr. aplikacije za spremljanje in analizo navigacijskega in poizvedovalnega vedenja uporabnikov, informacijski portali, digitalni tezavri, podatkovna skladišča*). Zapisano pomeni, da ni možno zgraditi kakovostno in učinkovito adaptivno digitalno knjižnico, ne da bi pri tem velikem podvigu med sabo tesno sodelovali vsaj bibliotekarji in informatiki/računalnikarji. V prihodnje, se v zvezi z digitalnimi knjižnicami obetajo velike spremembe. Spodbudil se naj bi še zlasti interdisciplinarni in celovit pristop pri dojemanju ter v naslednji stopnji izgradnji digitalnih knjižnic, ki jih je potrebno odločneje opredeliti kot informacijske sisteme za upravljanje z znanjem in storitvami, kajti knjižnice vsebujejo mnogo različnih potencialov, ki jih zaenkrat človeške družbe še niso uspele docela spoznati oziroma uvesti v delovne/poslovne procese različnih podjetij in zavodov (*npr. uvajanje različnih koristnih storitev kot so pomoč še zlasti poslovnim uporabnikom pri analizi podatkov, organizacija strokovnih znanj za poslovne uporabnike, analiza bibliografskega omrežja raziskovalcev z različnih področij znanosti in umetnosti, vsebinska analiza oziroma odkrivanje zakonitosti v besedilih znanstvenih*

*publikacij, razvijanje modelov in postopkov za lažje usvajanje učne ali študijske snovi – pomoč pri izdelavi didaktičnih pripomočkov, pomoč uporabnikom pri izdelavi delovnih informacijskih plošč ali armatur).*

Premalo je raziskav v zvezi s proučevanjem vedenjskih vzorcev uporabnikov oziroma ugotavljanja profilov uporabnikov na spletnih (*digitalnih*) knjižnicah, kar pomeni za kakovostno vzpostavitev in nadaljnjo razvijanje adaptivnih digitalnih knjižnic velik primanjkljaj, kajti uporabniki pomenijo nekakšen primarni ključ za delovanje sleherne uporabniške orientirane knjižnice.

## **1.2 Uporabniško usmerjen pristop k problematiki raziskovanja uporabnikov na digitalnih (spletnih) knjižnicah**

Pričujoče doktorsko delo predlaga uporabniško usmerjen pristop k reševanju problematike raziskovanja uporabnikov na digitalnih knjižnicah. Jedro pristopa je ogrodje za raziskovanje vedenjskih (*poizvedovalnih*) vzorcev uporabnikov, ki obravnava sociotehnične poglede z namenom izgradnje (*kvalitativnega modela*) adaptivne digitalne knjižnice s pomembnim modulom multidimenzionalnega rang tezavra. Ogrodje za raziskovanje vedenjskih vzorcev in v naslednji stopnji profilov uporabnikov sestavljajo naslednji gradniki:

### **→ Načrtovanje podatkov in priprava podatkov za ugotavljanje zakona moči in odkrivanje zakonitosti v podatkih (*Data Mining/Text Data oz. Text Data Mining*):**

Pri načrtovanju in pripravi podatkov za ugotavljanje zakona moči in odkrivanja zakonitosti v podatkih (*sem se vzgledoval*) po metodologiji CRISP, katera vključuje naslednje prvine: razumevanje področja, razumevanje podatkov, priprava podatkov, modeliranje podatkov, ovrednotenje podatkov in izgradnja podatkov. Na podlagi načrtovanih in pripravljenih podatkov, je bilo možno načrtovati in pripraviti podatke za ugotavljanje zakona moči in v nadaljnji stopnji odkrivati oziroma ekstrahirati nova znanja.

**→ Metoda ugotavljanje zakona moči (*angl.: Power Law*):** na podlagi zbranih webometričnih (*spletometričnih*) podatkov poizvedb uporabnikov na pilotni digitalni knjižnici in klasifikacije poizvedb; metoda ugotavljanja zakona moči omogoča dokazovanje univerzalnega obnašanja uporabnikov glede njihovih področnih interesov, ki se tudi na daljši časovni rok ne spreminja! Prav dokazovanje tovrstnega poizvedovalnega vedenjskega vzorca pomeni nekakšna odskočna platforma za kasnejše odkrivanje zakonitosti v podatkih oziroma v nadaljnji stopnji odkrivanja novih znanj.

→ **Odkrivanje zakonitosti v podatkih in besedilu:** na podlagi uporabe nekaterih metod s področja odkrivanja zakonitosti v podatkih in besedilu (*npr. metoda asociacijskih pravil, metoda hierarhičnih grozdov*), je možno odkrivati posebne vzorce in v nadaljnji stopnji nova znanja, kar nas v nadaljevanju pripelje do izgradnje kvalitativnega modela adaptivne digitalne knjižnice s pomembnim modulom oziroma prototipom multidimenzionalnega rang tezavra, ki je prilagojen uporabnikovim interesom.

→ **Izgradnja kvalitativnega modela adaptivne digitalne knjižnice z**

**multidimenzionalnim rang tezavrom:** izgradnja kvalitativnega modela adaptivne digitalne knjižnice z modulom multidimenzionalnega rang tezavra je izid vseh predhodnih prizadevanj in s tem posledično uporabniško usmerjenega pristopa k problematiki raziskovanja uporabnikov na digitalnih (*spletnih*) knjižnicah, kar v nadaljnji stopnji pripelje do vzpostavitve prave adaptivne digitalne knjižnice, ki se nenehno prilagaja informacijskim potrebam uporabnikov.

Predlagano ogrodje za raziskovanje vedenjskih poizvedovalnih idr. vzorcev uporabnikov na pilotni digitalni (*spletni*) knjižnici in v končni fazi razvijanja kvalitativnega modela adaptivne digitalne knjižnice omogoča izgradnjo prave kakovostne adaptivne digitalne knjižnice z multidimenzionalnim tezavrom, ki se prilagaja različnim uporabnikom in tako upošteva različne uporabniške poglede.

### 1.3 Zgradba doktorskega dela

Poleg uvodnega poglavja, sklepa, priloge, navedba virov in spremnega gradiva (*CD-ROM*) doktorsko delo vključuje še štiri ključna poglavja:

- **Poglavje 2: Teoretična izhodišča:** v tem poglavju so predstavljena teoretična izhodišča za različna področja: webometrija (*predstavitev Zipfovega zakona in uporaba njegove izpeljane različice* → zakon moči → *angl.: Power Law*), ontologija (*predstavitev področja ontologije in uporaba UDK klasifikacije*), tezaver (*opredelitev tezavra, zgradba tezavra, odnosi med pojmi, vrste tezavrov*), odkrivanje zakonitosti v podatkih (*opredelitev, predstavitev uporabljenih metod in tehnik*), digitalna knjižnica (*opredelitev, razvoj, adaptivna digitalna knjižnica idr.*), profiliranje uporabnikov (*na podlagi poizvedb na spletnem iskalniku idr.*) in spletni iskalniki (*razmišljanja o smislu uporabe spletnih iskalnikov*).

- **Poglavje 3: Analiza za pripravo modela:** pričujoče poglavje je bilo namenjeno predstavitvi analize za pripravo in zasnove modela ter v nadaljevanju za načrtovanje in pripravo podatkov (*vzgledoval sem se po metodologiji CRISP*), ki bi bili ustrezni za kasnejšo

odkrivanje zakonitosti v podatkih/besedilih (*klasifikacija poizvedb, izvedba zakona moči s preizkusi, priprava besedilnega dela podatkov, čiščenje poizvedb, opis podatkov, kratko poročilo o čiščenju podatkov, priprava podatkov s pomočjo programskega orodja Orange Canvas, priprava podatkov za ugotavljanje asociacijskih povezav in hierarhičnih grozdov*).

- **Poglavje 4: Odkrivanje zakonitosti v podatkih v okviru modela:** v najboljšejšem četrtem poglavju sem odkrival zanimive vzorce v podatkih, ugotavljal glavne značilnosti uporabnikov/poizvedovalcev, določil range posameznih UDK področij z ozirom na pomembnejše ključne besede, razvil metamodel okvirne miselne hierarhije interesnih področij uporabnikov (*oziroma spletnih obiskovalcev poizvedovalcev*) poizvedovalcev, uporabil metode in tehnike s področja odkrivanja zakonitosti v podatkih pred in po čiščenju podatkov, ovrednotil podatke, izpeljal nova spoznanja in v nadaljnji stopnji nova znanja o podatkih, dokazal prvo zastavljeno hipotezo in pri tem ovrgel drugo, nakar sem segmentiral uporabnike.

- **Poglavje 5: Kvalitativni model adaptivne digitalne knjižnice in ovrednotenje :** na podlagi predhodno pridobljenih novih znanj sem izgradil končni kvalitativni model adaptivne digitalne knjižnice s poudarkom na multidimenzionalnim rang tezavrom. Kvalitativni model sem na koncu še kritično ovrednotil.

## 1.4 Cilj doktorskega dela

V svoji doktorski disertaciji se bom na podlagi zbranih podatkov iz obdobja od 2003 do 2006 ukvarjal z analizo uporabe notranjega iskalnika s strani uporabnikov na obstoječi pilotni digitalni (*spletni*) knjižnici in s tem posledično z uporabo določenih webometričnih metod (*zakon moči – Power Law, analiza časovnih vrst*). Analiza uporabe iskalnika na svetovnem spletu je ena od področij webometrije (*spletometrije*), kajti ugotavljanje uporabe iskalnika s strani uporabnikov nam lahko daje veliko povratnih informacij o le-teh.<sup>1</sup> Na podlagi teh povratnih informacij bom v kasnejši stopnji z že omenjenimi metodami in drugimi (*kot npr. določene metode s področja odkrivanja zakonitosti v podatkih*) izdelal profile interesnih področij uporabnikov. Ugotovljeni profili mi bodo omogočali ustvarjati model adaptivne digitalne knjižnice, ki bo izpostavil pomen prilagojenosti in uporabnosti informacij na digitalni knjižnici z vidika uporabnika.

---

<sup>1</sup> Puntheeranurak, S. & Tsuji, H. (2005). Mining Web logs for a personalized recommender system. 3<sup>rd</sup> International Conferenc on Information Technology: Research and Education, 27-30 June, Hsinchu, Taiwan, 445-8

## 1.5 Opis problematike

Kompleksne družbene pojave težko oziroma že kar nemogoče razrešujemo samo s klasičnimi statističnimi metodami. Potrebno je dolga leta opazovati in zbirati ogromno količino podatkov, da bi lahko odkrili določene konstante (*npr. izobraževalni programi s strani Ministrstva za šolstvo*) in/ali vzorce, ki se odmikajo od konstantnih gibanj. Povrh tega je potrebno izbirati ustrezna orodja za analizo zbranih podatkov tako za besedne kot tudi številčne podatke, kajti izid uporabnikovih poizvedb je tako v obliki pojmov kot tudi frekvenc. Analiza poizvedb tako tudi zahteva pretvorbo v strukturirano obliko tj. v številčne podatke.

## 1.6 Pričakovani prispevki k znanosti

Na podlagi webometričnih in drugih analiz uporabe iskalnika na digitalni knjižnici bo možno izdelati model adaptivne digitalne knjižnice, ki izpostavlja pomen prilagojenosti in uporabnosti informacij za uporabnike, kar je še zlasti pomembno za podporo tako poslovnih, znanstvenoraziskovalnih in vzgojnoizobraževalnih procesov.<sup>2</sup> Obdelani podatki in v kasnejši stopnji pridobljena spoznanja o uporabnikih digitalne knjižnice se tako lahko posredujejo kompetentnim osebam (*npr. vodja oddelka, profesor*), ki zmorejo izboljšati oziroma reorganizirati delovne/poslovne procese v podjetjih ali zavodih.<sup>3</sup>

Osnovni doprinos doktorske disertacije k znanosti je izdelava modela izboljšane adaptivne digitalne knjižnice (*širše gledano: informacijski sistem za upravljanje z znanjem*), ki sem ga izdelal na podlagi socioloških (*spremljanje informacijskih potreb v vzgojnoizobraževalnem zavodu*) in webometričnih preučevanj lastne pilotne digitalne knjižnice (*v letih od 2001 do 2006*).<sup>4</sup> Model temelji na predpostavki o obstoju ustaljenih vedenjskih izobraževalnih vzorcih uporabnikov, ki sem jih dobil na podlagi dejavnosti le-teh in webometričnih analiz poizvedb. Izide sem v nadaljnji stopnji še obdelal z nekaterimi metodami s področja odkrivanja zakonitosti v podatkih (*npr. asociacijska pravila,*

---

<sup>2</sup> Jespen, E.T., Seiden, P., Ingversen, P. & Björneborn, L.(2004). Characteristics of scientific web publications: preliminary data gathering and analysis. *Journal of the American Society for Information Science and Technology*, 55(14): 1239 - 1249.

<sup>3</sup> Koohang, A.(2004). Students' perceptions toward the use of the digital library in weekly web-based distance learning assignments portion of a hybrid programme. *BRITISH JOURNAL OF EDUCATIONAL TECHNOLOGY*, 35 (5), 617-626.

<sup>4</sup> Jespen, E.T., Seiden, P., Ingversen, P. & Björneborn, L.(2004). Characteristics of scientific web publications: preliminary data gathering and analysis. *Journal of the American Society for Information Science and Technology*, 55(14): 1239 - 1249.

*hierarhično grozdenje*) z namenom, da izdelam interesne profile uporabnikov.<sup>5</sup> V kvalitativni model adaptivne digitalne knjižnice bom tako vključil še dodaten modul multidimenzionalnega tezavra, v katerem naj bi bile informacije organizirane na osnovi interesnih profilov uporabnikov, znotraj katerih so prikazani različni odnosi med deskriptorji (*hierarhija, ekvivalentnost, asociativnost, sinonimnost, antonimnost, homonimnost*). Obravnavani multidimenzionalni tezaver vključuje tri različne dimenzije, ki so med sabo tudi v ožji povezavi in so naslednje:

- program srednješolskega vzgojnoizobraževalnega procesa
  - profilirani uporabniki
  - rangirani deskriptorji, ki so izid poizvedb na notranjem iskalniku pilotne digitalne knjižnice
- Osnovni namen multidimenzionalnega tezavra je v tem, da uporabnikom omogoča bolj prilagojen in s tem posledično učinkovitejši dostop do uporabnih informacij (*zapisano velja tako za digitalne knjižnice na spletu kot tudi na CD-ROM*). Za indeksiranje vsebin s pomočjo obstoječega notranjega iskalnika, bi to pomenilo poenostavitev postopka indeksacije, gospodarnejšo/natančnejšo indeksacijo vsebin in s tem posledično nižjo stopnjo poizvedovalnega šuma po izvedenih poizvedbah. Povrh tega je možno multidimenzionalni tezaver uporabiti kot orodje pri razreševanju informacijskih problemov tako v informacijsko dokumentalističnih centrih kot tudi v knjižnicah (*to v sklopu referenčnega procesa bibliotekar/dokumentalist v odnosu z uporabnikom*).

**Izvirni prispevek k znanosti** je model adaptivne digitalne knjižnice z dodatnim modulom multidimenzionalnega tezavra (*še zlasti njegovih dimenzij*), ki ga bom v okviru svojega raziskovalnega dela predstavil tako, da bom:

- prikazal interesne profile uporabnikov in določil identičnosti ter različnosti med njimi
  - določil zgradbo multidimenzionalnega tezavra
  - kvalitativno prikazal povezave med dimenzijami šolski program, uporabnik in rangirani deskriptor (*izdelal primere za deskriptorje, vključujoče prikazov različnih odnosov med njimi*
- Predlagani prispevek k znanosti bom predstavil v obliki kvalitativnega modela, ki vključuje multidimenzionalni tezaver, ki ga bom prikazal kot prototipno izvedbo - prototipno izvedbo si bo možno ogledati na priloženem CD-ROM*).

---

<sup>5</sup> Koohang, A.(2004). Students' perceptions toward the use of the digital library in weekly web-based distance learning assignments portion of a hybrid programme. BRITISH JOURNAL OF EDUCATIONAL TECHNOLOGY, 35 (5), 617-626.

## 1.7 Metodologija in metodološka orodja

Metoda raziskovanja: popolno opazovanje naključne populacije ljudi oziroma spletnih obiskovalcev – v letih od 01.01. 2003 do 01.01.2006. Sproti sem spremljal dejavnosti poizvedovanja spletnih obiskovalcev na pilotni digitalni (*spletni*) knjižnici. Pri analizi zbranih podatkov bom v prvem delu doktorskega dela uporabil »Zakon moči« (*uporabil bom programsko orodje WinIdams in MS Excel*). V drugem delu doktorskega dela bom uporabil metode s področja odkrivanja zakonitosti v podatkih (*pri tem se bom posluževal naslednjih programskih orodij: Orange Canvas, CBA association Rules, Ilog Discovery in AntConc 3.1.302*).

## 1.8 Raziskovalne hipoteze in vprašanja

Na podlagi razvrščenih poizvedb sem izpeljal naslednje raziskovalne hipoteze in raziskovalna vprašanja.<sup>6</sup>

### 1.8.1 Raziskovalne hipoteze

- 1.) Informacijski/poizvedovalni vedenjski vzorec spletnih obiskovalcev na obstoječi pilotni digitalni knjižnici, se dolgoročno gledano ne spreminja in je tako neodvisen od različnih časovnih obdobj.
- 2.) Informacijski/poizvedovalni vedenjski vzorec spletnih obiskovalcev na obstoječi pilotni digitalni knjižnici je zelo odvisen od različnih časovnih obdobj in se tako nenehno spreminja.

### 1.8.2 Raziskovalna vprašanja

- 1.) Kakšne izide (*spoznanja*) lahko pričakujem glede na uporabljene metode in orodja?
- 2.) Ali so te metode ustrezne?
- 3.) Ali se informacijski/poizvedovalni vedenjski vzorec spletnih obiskovalcev na obstoječi pilotni digitalni knjižnici dolgoročno ne spreminja?
- 4.) Ali se informacijski/poizvedovalni vedenjski vzorec spletnih obiskovalcev na obstoječi pilotni digitalni knjižnici skozi čas nenehno spreminja?
- 5.) Ali je možno profilirati uporabnike na podlagi uporabljenih metod in orodij?

---

<sup>6</sup> Zamisel se mi je porodila na podlagi naslednjega vira: Björneborn, L. & Ingversen, P.(2004). Toward a basic framework for webometrics. *Journal of the American Society for Information Science and Technology*, 55(14): 1216-1227.

- 6.) Kaj storiti v primeru, če uporabljene metode niso ustrezne in kako se bo to odražalo na izdelavo modela adaptivne digitalne knjižnice?
- 7.) Katere metode bo potrebno uporabiti, če zakon moči dokazuje, da so informacijski/poizvedovalni vedenjski vzorci spletnih obiskovalcev zelo dinamični in spremenljivi?
- 8.) Kateri informacijski/poizvedovalni vzorci spletnih obiskovalcev so zelo konstantni?

## 2 Teoretična izhodišča

V tem poglavju bom predstavil področje webometrije (*spletometrija*), ontologije (*še zlasti univerzalni decimalni klasifikacijski sistem in*) in način razvrščanja poizvedb, ki so jih spletni obiskovalci izvedli (*od 01.01.2003 do 01.01.2006*) na notranjem iskalniku lastne digitalne (*spletne*) knjižnice, določene pojme in metode oziroma tehnike s področja (*spletnega*) odkrivanja zakonitosti v podatkih, nakar bom na kratko poročal še o digitalnih knjižnicah. Poglavje o teoretičnih izhodiščih bom zaključil s kratkim opisom profiliranja uporabnikov in smotra uporabe spletnih iskalnikov.

### 2.1 Webometrija (*spletometrija*)

Pojem WEBOMETRIJE (*v nadaljevanju spletometrija*) je v znanstvenem svetu sorazmerno nov, čeprav izhajajo prve zamisli o možnosti nastanka tovrstne informacijske vede (*nekateri avtorji spletometrijo poimenujejo tudi kot metodo*) približno iz leta 1995/1996. Spletometrija je v bistvu nastala kot posebna veja s področja informetrije, scientometrije oziroma še zlasti bibliometrije.<sup>7</sup>

Osnovna naloga spletometrije (*širše: Cybermetrics se nanaša na celotni Internet oz. sin. Medmrežje in ne zgolj na svetovni splet oz. World Wide Web*) je meriti in v naslednji stopnji ovrednotiti tako spletne publikacije (*spletne knjige, spletne revije, spletne članke ipd.*), ustvarjalne in manj ustvarjalne spletne avtorje in posamezna znanstvena področja (*npr. dejavnik kakovosti spletnih prispevkov s področja književnosti, tehnike, medicine itd.*).<sup>8</sup> Komercialna oblika spletometrije nam je že nekoliko bolj znana, saj posega na področje merjenja in ovrednotenja povpraševanj spletnih obiskovalcev po posameznih izdelkih, ki se ponujajo po t.i. spletnih trgovinah. Pravi začetni razcvet za znanstveno sfero pa je spletometrija dosegla šele s prihodom 21. stoletja (*npr. 2001 Danec Lennart Björneborn, Peter Ingversen – idejni oče pojma webometrije in predlagatelj pojma Web Impac Factor, Mike Thelwall, Liwen Vaughan, 2004 Nemeč Phillipp Mayr*).<sup>9</sup> Spletometrijo je možno pojmovati še širše in to v povezavi z informacijsko znanostjo, komunikologije, sociologije, psihologije, statistične fizike idr. Spletometrijo je možno tudi povezati z Data Mining-om

---

<sup>7</sup> Mayr, P.(2004). Entwicklung und Test einer logfilebasierten Metrik zur Analyse von Website entries am Beispiel einer akademischen Universitäts-Website. Berlin : Inst. für Bibliothekswiss, 106 str. (Humboldt-Universität zu Berlin, Institut für Bibliothekswissenschaft) na strani 10.

<sup>8</sup> Thelwall, M. & Vaughan, L.(2004). Webometrics: An introduction to the special issue. Journal of the American Society for Information Science and Technology, 55(14): 1213 - 1215.

<sup>9</sup> Björneborn, L. & Ingversen, P.(2001). Perspectives of webometrics. Scientometrics, 50(1): 65-82.

(odkrivanje zakonitosti v podatkih) oziroma z njegovim podpodročjem Web Mining-om (odkrivanje zakonitosti v podatkih na spletu).

V spletometriji so razvili tudi modelirno tehniko za prikazovanje spletnih povezav in glavnih spletnih vozlišč (*Consistent Web node Framework*), ki vsebuje naslednje glavne gradnike:<sup>10</sup>

- krog: ponazarja glavne spletne strani
- pravokotnik: manjše oziroma posamične spletne strani
- diagonalna črta: spletni direktoriji
- trikotnik: ponazarja države, domene idr.

Spletometrija se v glavnem največ ukvarja z naslednjimi analizami:

- 1.) Analiza vsebine spletnih strani (*npr. informacijska ponudba na digitalni knjižnici*)
- 2.) Analiza zgradbe spletnih povezav in/ali spletnih strani
- 3.) Analiza uporabe svetovnega spleta (*dnevnik spletnega strežnika, iskanje oziroma iskalne tehnike spletnih obiskovalcev idr.*)
- 4.) Analiza spletne tehnologije (*vkjučuje med drugim tudi delovanje spletnih aplikacij ipd.*)

Pri tovrstnih štirih analizah se uporabljajo različne metode, ki se raztezajo od analize časovnih vrst pa tja do merjenja in računanja zakona moči (*Power Law – Zipfov zakon, Bradfordov zakon, Lotkov zakon ipd.*). Še zlasti o Zipfovem zakonu bom v tem teoretičnem uvodu nekoliko obširneje poročal, ker nameravam v drugem poglavju uprabititi njegovo izpeljano različico. Preden bralcem tega dela predstavim omenjeno snov, bi jih želel še opozoriti na naslednjo zbirko spletnih povezav <http://www.beepworld.de/members61/soc2>, kjer si bodo lahko bralci tega dela lahko ogledali (*v primeru, da si to želijo*) različne možnosti in meritve na področju webometrije/spletometrije (*v teh primerih gre tudi za hibridne pristope*). Tovrstna gradiva bodo tudi dosegljiva v spremnem gradivu tega dela na CD-ROM, poleg meritev in nekaterih drugih poskusnih slikovnih prikazov ipd.

### **2.1.1 Zakon moči (*Power Law*)**

V informacijski znanosti poznamo različne zakone moči (*angl. Power Laws*), ki jih uporabljajo tudi druge znanosti in vede od fizike, kemije, matematike, jezikoslovja, bibliotekarstva oziroma bibliometrije, statistike pa tja do kot že omenjeno spletometrije. Tako v informacijski znanosti lahko poročamo o Zipfovem zakonu (*ta je mimogrede povedano tudi najstarejši in najpogosteje uporabljen*), Bradfordovem zakonu, Lotkovem zakonu (*ta je po*

---

<sup>10</sup> Björneborn, L.(2004). Small-World Link Structures across an Academic Web Space: A Library and Information Science Approach na strani od 11 do 16.

*mnenjih mnogih nekoliko manj zanesljiv in služi bolj kot nekakšna ocena)* idr. O slednjih dveh kot tudi o vseh ostalih manj uporabljenih v tem delu ne nameravam pisati. Izidi s pomočjo metod zakonov moči nam lahko pokažejo neko univerzalno obnašanje določenih pojavov tako v naravi kot tudi v družbi oziroma kompleksnih sistemih, kar nam ljudem lahko olajša njihovo razumevanje.

## 2.1.2 Zipfov zakon<sup>11</sup>

Profesor George Kingsley Zipf je še zlasti v letih od 1932 do 1949 izdal zelo pomembna in odmevna dela, med katerimi naj omenim dve: “Human Behavior and the Principle of Least Effort“ (v slov. *Človeško vedenje in princip najmanjšega odpora*) in “The Law of the Frequency Distribution of Words” (v slov. *Zakon o frekvenčni distribuciji korenov besed*). Njegovi temeljni postavki sta bili informativnost (*pridobivanje čim več koristnih informacij*) in udobje (*izključevanje nepotrebnih akcij, pri čemer naj bi si človek prihranil čas in energijo*). Ukvarjal se je med drugim tudi s populacijo velemest (*kot tudi s številom oziroma pogostosti uporabljenih besed v besedilih npr. Ulysses – James Joyce, biblija idr.*) in pri tem ugotovil, da obstaja med številom prebivalcev nekega mesta in njegovim rangom potenčna zveza,

$$N = C/(n - 2/5)^{3/4}$$

kjer je  $N$  število prebivalcev,  $n$  je rang mesta,  $C$  pa je konstanta, ki jo izračunamo po naslednjem obrazcu:  $C = N$  (*frekvenca*)  $\cdot$   $r$  (*rang*). To pomeni, če poznamo rang nekega poljubnega velemesta lahko ocenimo število prebivalstva tega mesta. Prav na ta izračun oziroma na njegovo izpeljano različico se nameravam v poznejšem pisanju še povrniti.

### 2.1.2.1 Zakon o frekvenčni distribuciji korenov besed

Luhn in Zipf sta besede iz angleškega jezika razčlenila po frekvencah in jih razvrstila po rangirnih vrstah. Pri tem sta prispela do ugotovitve, da je možno množično besed v grobem razdeliti v tri skupine:

- 1.) skupina – besede, ki se zelo pogostokrat pojavljajo in bistveno ne prispevajo k vsebini besedila.
- 2.) skupina – besede, ki se zelo redko pojavljajo in prav tako bistveno ne prispevajo k vsebini besedila.

---

<sup>11</sup> [http://www.mf.uni-lj.si/~jure/pred\\_bib](http://www.mf.uni-lj.si/~jure/pred_bib) (2003-12-04).

3.) skupina – ali vmesna skupina, katere besede so predstavnice vsebine besedila.

Na podlagi tega sta sklepala, da je možno na podlagi poznavanja frekvence in ranga besed določiti besede, ki pomenijo jedro vsebine (*odvisnost v obliki hiperbolične funkcije*). To je pomenilo, da imajo vsa daljša besedila neko značilno univerzalno zgradbo. Prav ta spoznanja so uporabili za avtomatično indeksiranje predstavnikov besedil v določeni računalniški zbirki (*metoda blokiranje, krnjenje, računanje povedne moči*).

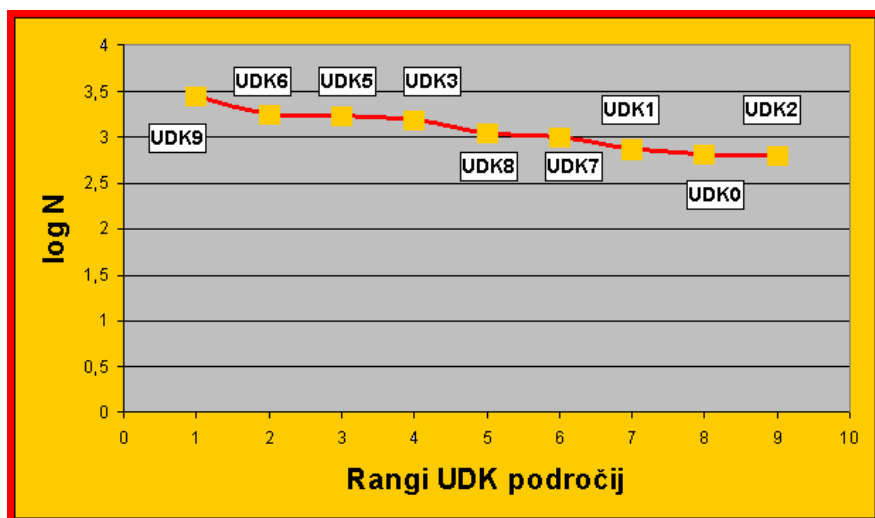
Ena od izpeljanih različic Zipfovega zakona je takšne narave, da logaritmiramo (*desetiški logaritem*) frekvence opazovanih množic, nakar jih primerjamo z ustreznimi rangi.<sup>12</sup>

$$C = \log N \cdot r$$

S tem pristopom dobimo kot izid bolj ali manj ravno premico, ki nam ponazarja neko sorazmerno konstantno dogajanje, ki je neodvisno od časa. Naj zaradi večje nazornosti podam testne podatke v preglednici in odgovarjajoč slikovni prikaz:

### 2.1.2.2 Preglednica 1: Zakon moči pri poizvedbah (2002 do 2005)

UDK 9	UDK 6	UDK 5	UDK 3	UDK 8	UDK 7	UDK 1	UDK 0	UDK 2
3,4551	3,2392	3,2291	3,1858	3,0484	3,004	2,8627	2,8122	2,7945



### 2.1.2.3 Slikovni prikaz 1: Primerjava logaritmiranih frekvenc z rangi

Preglednica in slika prikazujeta zakon moči glede poizvedb, ki so jih izvedli spletni obiskovalci na pilotni digitalni knjižnici po posameznih UDK (*univerzalna decimalna klasifikacija*) področjih za obdobje od 2002 do 2005. Izobraževalni interesi spletnih obiskovalcev so na podlagi dobljenih meritev in izidov zelo značilni in stabilni, saj ne kažejo

<sup>12</sup> Za natančnejše podatke o tej metodi gl. naslednje delo: Bak, P.(1997). How nature works : The Science of Self Organized Criticality. Oxford ; Melbourne ; Tokio : Oxford University Press, 212 strani. Delo je dosegljivo v Centralni tehniški knjižnici v Ljubljani.

pretirane odvisnosti od časa, kar pomeni, da se tovrstno obnašanje tudi na daljši časovni rok ne bo bistveno spremenilo. **Spletni obiskovalci torej kažejo na neko univerzalno obnašanje glede izobraževalnih interesov (ožje gledano) ali širše gledano glede področnih interesov (npr. za poslovanje, za zdravje, za organizacijo dela, za izobraževanje, za zabavo itd.), ki se tudi na daljši rok ne spreminja! Trditev velja za spletno izvedbo UDK leksikona, ki je v svoji vsebinski zasnovi zastavljen tako, da je zelo splošen in zelo širokega razpona.**

Več kot imamo meritev oziroma podatkov na voljo, toliko bolj je tudi zanesljiva ta metoda (od 5000 podatkov in več).

Podpoglavje o spletometriji je tako zaključeno. V naslednjem podpoglavju nameravam nekaj besed nameniti ontologijam, v katera uvrščamo tudi klasifikacijske sisteme, med katere prištejemo tudi univerzalno decimalno klasifikacijo (v nadaljevanju UDK).

## 2.2 Ontologija

Ontologija v izvornem pomenu je filozofska veda oziroma disciplina (grško: *ontos* = *bitje*; *logos* = *veda*). Ontologija pomeni nauk o biti, ki se ukvarja s temelji resničnosti in njenimi najsplošnejšimi lastnostmi kot so npr. snov, bistvo, pomen, osnovno počelo, prvi vzrok idr.<sup>13</sup> V informatiki je pomen ontologije nekoliko drugačen oziroma določnejši. Ontologija z vidika informatike je disciplina, ki se ukvarja z opredelitvijo besed in njihovim pomenom, kar se v nadaljnji stopnji uporablja za opis in predstavitev določenega znanja z bolj ali manj širokega področja. Ontologija v informatiki vsebuje pojme kot so razred, primerek, razmerje, lastnost, vrednost in pravilo.<sup>14</sup>

Pod ontologije (poleg semantičnih omrežij - *tezavri*, *predikatne logike* idr.) uvrstimo tudi taksonomije (npr. *klasifikacijski sistemi: od Fasetne, Cutterjeve, Blizzove, Kongresne knjižnične pa tja do Deweyjeve in univerzalne decimalne klasifikacije*). Taksonomija pomeni razvrstitev prvin v določen klasifikacijski sistem, v katerem so tudi bolj ali manj natančno ponazorjene določene povezave, kar je odvisno od uporabljenega klasifikacijskega sistema. V tem teoretičnem uvodu bom na kratko predstavil UDK, kajti ta klasifikacijski sistem sem uporabil za razrščanje poizvedb po glavnih razredih, ki so jih v obdobju od 01.01.2003 do 01.01.2006 izvedli spletni obiskovalci na notranjem iskalniku pilotne digitalne knjižnice (lastno izdelana spletna aplikacija, kjer so podatki razvrščeni po UDK – pozneje sem to

---

<sup>13</sup> Opredelitev povzeta po delu: Sruk, V.(1980). Filozofsko izrazje in repertorij. Murska Sobota: Pomurska založba, na str. Od 246 – 248.

<sup>14</sup> Gl. članek: Lavbič, D & Krisper, M. Semantika podatkov in ontologije iz leta 2005 na str. 5 – 7.

izvedbo še razširil na spletno knjižnico kot glavni modul in za logistiko, katerih meritve bodo prav tako dosegljive na CD-ROM).

### 2.2.1 UDK

Univerzalna decimalna klasifikacija je ena od najbolj razširjenih in uveljavljenih klasifikacijskih sistemov v svetu (v nadaljevanju UDK) in jo je možno uporabiti na številnih področjih (npr. v znanosti, umetnosti, industriji itd.). Bralcem tega dela je po vsej verjetnosti najbolj znana uporaba UDK v knjižnicah s prostim pristopom do gradiva, kjer razna knjižnična gradiva urejujejo po sistemu UDK. Nekoliko manj znana je uporaba UDK v vesoljski tehniki, v državnih upravah, v računalništvu (npr. ontologija, avtomatična UDK spletnih dokumentov) in nenazadnje v znanostih pri razvrščanju družboslovnih ali naravoslovnih pojavov v bolj ali manj podrobne razrede.

V tem delu bom prikazal možnost uporabe UDK za razvrščanje različnih poizvedb, ki so jih izvedli spletni obiskovalci na pilotni digitalni knjižnici po posameznih UDK področjih. Pri tovrstnih prizadevanjih, bi bilo možno uporabiti druge klasifikacijske sisteme, vendar menim, da je UDK eden najbolj razvitih klasifikacijskih sistemov nasploh. Naj zato na kratko predstavim ta klasifikacijski sistem.<sup>15</sup> V ta namen bom podal zelo kratko in kljub temu dovolj nazorno opredelitev s svoje seminarske naloge na predhodnem podiplomskem magistrskem študiju.<sup>16</sup>

"Čisto na kratko povedano je UDK klasifikacijski sistem, ki razvršča "vso" človekovo znanje, dejavnosti in materiale itd." s pomočjo števil (pravilneje UDK vrstilcev), črk, drugih znakov tj. matematičnih (npr. +, =, -) in interpunkcijskih znamenj (npr. :, ., ;). Prav ta prijetna lastnost UDK omogoča, da lahko zajamemo zelo široko različne podatke/informacije z različnih področij človekovega življenja, ki se raztezajo od naravoslovja, tehnike/aplikativnih ved, družboslovja, humanistike, umetnosti, športa pa tja do čisto vsakdanjih opravkov (npr. nakupovanje v delikatesni trgovini, igranje šaha, gledanje televizije, delo z računalnikom, branje knjige). Ta UDK pa je zapisan v tablicah, ki se členijo na glavne in pomožne, ter v katerih so zapisani UDK vrstilci kot tudi pravila." Prav ta pravila v glavnih in pomožnih tablicah narekujejo način povezovanja oziroma odnose (nadrejenost, podrejenost, enakovrednost idr.) med različnimi področij znanosti, izvedb, dejavnosti in umetnosti. Z ozirom na zastavljeno vsebino tega dela, bom natančneje opisal globalni pomen

<sup>15</sup> Bralci tega dela si lahko ogledajo natančnejšo obrazložitev UDK v priloženem CD-ROM ali pa na medmrežju na naslednjih URL-jih: <http://www.beepworld.de/members56/udkmmis/udk.htm> , <http://www.beepworld.de/members29/udkleksikon/uvod.htm> , <http://udkbravehost.com/legenda.htm> itd.

<sup>16</sup> Gl. URL: <http://udkbravehost.com/cps.htm> ali priloženi CD-ROM

UDK področij in pravila globalnega razvrščanja poizvedb po UDK in pri tem izpostavil posebnosti oziroma razlike nasproti uveljavljenega načina razvrščanja podatkov/informacij. Prav ta opis bo pomemben zaradi načrtovanja in priprave podatkov za uporabo zakona moči in za poznejše uporabljene metode odkrivanja zakonitostih v podatkih ter zaradi izpeljave logičnega konceptualnega klasifikacijskega modela, ki ga bom lahko v kasnejši stopnji uporabil, še zlasti za segmentacijo ali razčlenbo oziroma izdelavo profilov spletnih obiskovalcev/uporabnikov v posamezne miselne skupine kot tudi za odkrivanje povezav (*asociacij*) med različnimi poizvedbami oziroma skupinami poizvedb.

### **2.2.2 Globalni pomen UDK področij (*izpeljane osnovne logične značilnosti*)**

a.) UDK 0 - imenujemo tudi splošno področje. V to skupino razvrščamo vso človekovo znanje (*znanost, umetnost itd.*) najprej s sistematičnega (*npr. razvejenost znanosti, sistematika znanosti ipd.*) in z metodološkega vidika (*npr. metodološka orodja, metodologija*). Ob nadaljnji delitvi znanja prihaja do stika z naslednjimi prvinami znanja kot npr. razni pisni izdelki, vendar gledamo nanje povsem splošno (*diplomska dela, seminarske naloge, disertacije itd.*). V to skupino razvrščam tudi vso (*na splošno*) človekovo znanje iz preteklosti (*npr. muzeji, galerije, pismenstvo, pisave*) in sedanjosti (*npr. knjižnice, katalogi, informacijski sistemi, podatkovne baze, računalniki*) s splošnega vidika. V to skupino tako tudi razvrščam posebne enciklopedije, leksikone, slovarje kataloge ipd. in ne zgolj splošne kot se to izvaja pri izvornem oziroma dogovorjenem sistemu UDK (*npr. leksikon medicine gre v tem primeru pod globalno skupino UDK 0, sicer bi šlo pod 6(031)*). O tovrstnih posebnostih se kanim povrniti nekoliko pozneje!

b.) UDK 1 - imenujemo tudi področje filozofije, psihologije, logike in etike. V to skupino razvrščamo človekovo znanje, ki se osredotoča na bistvo in pomen človeka, še zlasti s poudarkom na posamezniku, ki je zahvaljujoče svojemu psihološkemu ustroju (*miselni procesi, čustvovanje, spominjanje ipd.*) v nenehnem stiku z naravo in z družbo. Prav ta stik pa človeku neobhodno predpisuje nek logičen in kakovosten odnos do okolja in do samega sebe (*etika, morala ipd.*). Prav etika in morala pomenita v klasifikacijskem pogledu nekakšen mehek prehod na UDK 2.

c.) UDK 2 - imenujemo tudi področje religije ali verstev (*krščanska in nekrščanska verstva*). Pozitivna verovanja v Boga ali višjo silo se ukvarjajo z obstojem človeka v odnosu do

vsemogočnega (*monoteistična verstva*) Boga in/ali do vsemogočnih bogov (*politeizem*). Ta odnos zahteva ustrezno čustveno podoživljeno etiko in moralo, ki povzdiguje racionalno oziroma naravno usmerjenega človeka v neko višjo poslanstvo do nekega višjega sveta. Šele tovrstna etika in morala omogočata pravilen pozitiven odnos do vsega kar obstaja ter v človeku vdahne neko dušo, ki spoštuje in stopa v stik, tako z bivanji na nižjih/manjših nivojih in/ali z bivanjih, ki nas ljudi prekašajo. Po mojem osebnem mnenju, bi morala biti ta UDK 2 skupina z razvojnega vidika človeštva označena kot prva skupina nasploh, kajti če se spomnimo primitivnih verovanj in stenskih risb praljudi, potem spoznamo, da se je znanost razvijala precej pozneje!

d.) UDK 3 - to je področje družboslovnih ved (*sociologija, ekonomija, pravo, politika, izobraževanje, sociala, meroslovje, etnologija itd.*). Šele, ko je vsebina človeka osmišljena z bistvom, etiko, moralo, logiko, z višjimi miselnimi procesi in s pravo pozitivno vero in/ali ideologijo, se lahko posamezniki med seboj povezujejo v družbo/-e. Vse, kar je z vidika človekovega obstoja in dejavnosti družbenega predznaka, se uvršča v to skupino UDK 3.

e.) UDK 4 - po mednarodnem dogovoru je ta skupina sicer prazna. V to skupino razvrščam poizvedbe, ki jih ni možno razvrstiti v vse ostale UDK skupine (*npr. težko prepoznavne besedne igre, uporaba nenavadnih kratic, nove oziroma nepoznane vede*). Tovrstnih poizvedb v bistvu ni prav mnogo.

f.) UDK 5 - ali področje naravoslovnih ved (*ekologija, matematika, astronomija, fizika, kemija, paleontologija, geologija, meteorologija, biologija, zoološka oziroma fizična antropologija itd.*). Vso našo izvorno naravno okolje, ki ga kot takšnega prepoznamo in priznamo ter s katerim smo v nenehnem stiku, se večinoma nahaja izven našega človeškega telesa (*izjeme so lahko npr. virusi, bakterije, paraziti - mikrosvet*). Ta svet učinkuje oziroma vpliva na nas ljudi v zelo veliki meri, čeprav vrši človeška vrsta v vsej svoji evoluciji precejšen vpliv na naravo, ki pa ni vedno bil zelo etičen in moralen. Ta UDK 5 skupina v bistvu pomeni tudi nekakšen prehod na naslednjo UDK 6 skupino, kajti v UDK 5 skupini se zbirajo v veliki meri teoretična spoznanja in z eksperimenti pridobljeni dokazi, ki se v nadaljnji stopnji praktično uporabljajo.

g.) UDK 6 - področje medicine in uporabnih ved (*medicina, inženirske vede: npr. strojništvo, elektrotehnika, naprave, podjetja, gradbeništvo, predelani materiali, kemične industrije*). Vse

kar je v povezavi z zgradbo človeka in njegovim psihološkim ter še zlasti fiziološkim zdravjem lahko uvrstimo v UDK 6. Povrhu tega razvrščamo v to skupino vso človekovo izvedbeno znanje (*praktična znanja*).

h.) UDK 7 - področje umetnosti in umetnostne kulture (*teorija umetnosti, estetika, arhitektura, grafika, slikarstvo, kiparstvo, zbirateljstvo, glasba, šport, turizem, zabave in igre ipd.*). To področje se lahko zelo krepko navezuje na filozofijo, sociologijo in verstvo. Z razvojnega vidika človeštva, bi lahko bila ta skupina razvrščena v isto skupino kot verovanja. V to skupino se razvrščajo človekova avdio- vizualna izrazila (*npr. glasba, slikarstvo*), vizualna- tipna izrazila (*npr. kiparstvo*), vizualna- gibalna in/ali zvočna izrazila (*npr. nogomet, ples, balet, igre*) in vizualna- prostorska izrazila (*npr. arhitektura*). Vsebinsko gledano pomeni to področje nekakšen spoj med estetiko (*razmerje med čutnostjo, čustvi in razumom*) in razumom (*razmerje med mislijo in izražanjem in/ali udejanjanjem misli*).

i.) UDK 8 - področje jezikoslovja in književnosti (*jezikoslovje, fonetika, morfologija, dialektologija, književnost, literarna veda, zvrsti, literarna zgodovina itd.*). V ospredju so besede in njihove kombinacije, raziskovanje besed, sporočanja in sporazumevanje v različnih jezikih. V prvem predelu UDK 8 se preučujejo in vzpostavljajo zakonitosti in v naslednji stopnji se ustvarjajo pravila za besedno sporazumevanje tako v vsakdanjem življenju kot tudi na različnih delovnih mestih (*npr. znanstveni jezik, umetniško sporazumevanje, sporazumevanje v podjetjih*). V drugem predelu se natančneje razčlenjuje in prikazuje področje književnosti, ki pretežno z besedami prikazuje življenje (*in številne situacije*), ki je nekoč bilo, je sedaj in ki bi lahko bilo. V bistvu predstavlja UDK 8 nekakšno ogrodje za vse družbene odnose, kajti brez jezika, bi se ljudje težje povezovali in s tem posledično bi težje prišlo do vzpostavitve različnih človekovih družb.

j.) UDK 9 - področje razvoja sveta in človeka (*arheologija, domoznanstvo, geografija, življenjepisi/raziskave pretežno znamenitih ljudi in zgodovina*). Z UDK vrstilci se opisuje razvoj sveta in človeka skozi različne dogodke oziroma obdobja, ki so bili najbolj izraziti tako po moči kot tudi po vztrajnosti (*semkaj lahko prištejem tektonske premike oziroma preobrazbo sveta, svetovne vojne, znamenite oziroma vplivne osebnosti, rodbine; skratka v to skupino se uvrstijo vsi dejavniki, ki so izrazito vplivali na svet in s tem na človeštvo zaradi delovanja naravnih in družbenih sil*).

Naj zdaj v nadaljevanju navedem pravila globalnega razvrščanja.

### **2.2.3 Pravila globalnega razvrščanja poizvedb po UDK**

Globalno razvrščanje ali klasificiranje poizvedb na notranjem iskalniku se izvede na podlagi dogovorjenih pravil sistema UDK-ja, na podlagi gibljivih pravil (*ob tem dejavno spremljam premike spletnih obiskovalcev po posameznih spletnih straneh po izvedenih poizvedbah*) in po utežeh (*s pomočjo programskega orodja PICO SEARCH štejem tudi število določenih poizvedb, ki jih je nek spletni obiskovalec izvedel*). Dogovorjena pravila so izhodišče in imajo prednost pred gibljivimi. Upoštevanje uteži, ki pomeni število ali frekvenco določene poizvedbe pa dopolnjuje in olajšuje globalno UDK razvrščanje izvedenih poizvedb spletnih obiskovalcev. To pomeni, da so uteži podrejene dogovorjenimi in gibljivimi pravili. Za globalno UDK razvrščanje poizvedb se uporabljajo naslednji pripomočki: UDK glavne tablice (*splošni in posebni privesni vrstilci se pri tem razvrščanju uvrstijo v ustrezne glavne UDK skupine*), UDK geselski katalog, COBISS/COBIB, spletni slovar pojmov, spletni slovar Slovenskega knjižnega jezika in različni iskalniki (*za razreševanje tujk ali nepoznanih pojmov in za določevanje glavnih UDK vrstilcev*).

Gre za sintezo treh kazalcev tj. razvrščanje poizvedbe po UDK, razvrščanje poizvedb z ozirom na UDK in na premik spletnih obiskovalcev po straneh UDK leksikona ter nenazadnje upoštevanje uteži oziroma število določene poizvedbe, ki jo je napravil spletni obiskovalec. V mnogih primerih globalnega razvrščanja poizvedb po UDK ni potrebno uporabljati vse tri kazalce, ker so primeri poizvedb dovolj nedvoumne. V nekaterih primerih globalnega razvrščanja je potrebno uporabiti prav vse tri kazalce. O tem bom pisal v nadaljevanju in bom podal nekaj primerov iz prakse.

### **2.2.4 Nekaj primerov globalnega UDK razvrščanja na podlagi spletometričnih podatkov iz prakse**

V tem podpoglavju bom sistematično in globalno po posameznih UDK področjih navedel nekaj nedvoumnih ter mejnih primerov razvrščanja poizvedb, ki so jih spletni obiskovalci izvedli na spletnem UDK leksikonu.

1. UDK 0 - npr. poizvedba leksikon se razvršča v vsakem primeru kot UDK 0. IZJEMA: Pri tem globalnem UDK razvrščanju poizvedb sem vzpostavil pravilo, da se vsi leksikoni, vse enciklopedije, slovarji, katalogi ipd. razvrščajo v UDK 0, pa najsi je govora o splošnih ali pa o posebnih leksikonih ipd. V to področje sicer razvrščam vse poizvedbe, ki so splošne narave in

ne natančno opredeljujejo posebnosti. Tako npr. poizvedbo "Organizacija in ravnateljstvo razvrščam kot UDK 0", medtem ko se pa poizvedba "Organizacija in ravnateljstvo v industriji" razvršča pod UDK 6, natančneje pod uporabne vede. Pri poizvedbah, ki opredeljujejo pisne izdelke npr. "Seminarska naloga" postopam podobno. Iskalno geslo "Seminarska naloga" razvrščam pod UDK 0, medtem ko "Seminarska naloga medicina" razvrščam pod UDK 6. Drugi primeri: "Seminarska naloga kemija" razvrščam pod UDK 5; "Diplomsko delo ekonomija" gre pod UDK 3, a poizvedbo "Diplomsko delo" razvrščam pod UDK 0; Metode in metodologijo na splošno razvrščam pod UDK 0, toda npr. metode posebnega znanstvenega področja npr. fizikalne metode pod UDK 5 itd.

2. UDK 1 - npr. poizvedba "Aristoteles o razlaganju" razvrščam pod UDK 1, ker gre v tem primeru za pridobitev informacije o Aristotelovem delu, ne pa o njegovem življenju. V primeru, da bi se poizvedba glasila zgolj "Aristoteles", bi to poizvedbo vedno razvrstil pod UDK 9, kajti osebe, rodbine, države, svetovni zgodovinski dogodki, domoznanstvo, arheologijo se vedno razvrščajo pod UDK 9. Velja omeniti mejne ali interdisciplinarne primere poizvedb kot npr. "Filozofija kemije". V tem primeru tovrstno poizvedbo razvrstim vedno pod UDK 1, razen za naslednja primera:

a. Ko ugotovim, da je spletni obiskovalec izvedel več poizvedb npr. "Filozofija kemije" je izvedel enkrat, nato "Kemija" dvakrat in da je povrh tega spletni obiskovalec obiskal zgolj področje UDK 5, potem razvrščam vse tovrstne ipd. poizvedbe pod UDK 5, vključno s poizvedbo "Filozofija kemije"! Pri tem razvrščanju sta mi pomagala še ostala dva kazalca kot sta premik spletnega obiskovalca (*ogledal si je področje UDK 5. Tovrstne premike ugotavljam s spletnima programskima orodjema kot sta Nedstat in Extreme Tracking*) in uteži posameznih poizvedb (*prvo poizvedbo je spletni obiskovalec izvedel enkrat, drugo poizvedbo pa dvakrat. Pri tem dobimo razmerje 1(za UDK 1) :2(za UDK 5)*).

b. Spletni obiskovalec je izvedel poizvedbo "Psihologija družbe" enkrat, a poizvedbo "Sociologija" trikrat. Poizvedbo "Psihologija družbe" razvrščam pod UDK 1 in poizvedbo "Sociologijo" pod UDK 3. V tem primeru nastane zopet enostavna računica, ki se glasi: razmerje poizvedb 1(za UDK 1):3(za UDK 3), s katere lahko izpeljem naslednjo razvrščanje: Poizvedba "Psihologija družbe" je interdisciplinarne narave in se natančneje po UDK - ju razvršča kot 159.9:316 ali pa kot 316:159.9, kar je odvisno od vsebinskega poudarka, v tem primeru spletnega obiskovalca. S pomočjo uteži ali številčnosti (*frekvence*) določene

poizvedbe lahko v tem primeru določim vsebinski poudarek, ki gre v prid UDK 3, kar pomeni, da se vse poizvedbe razvrščajo kot UDK 3. V tem primeru ni bilo potrebno spremljati premike spletnega obiskovalca. Naj ponovno poudarim, da so tovrstne poizvedbe spletnih obiskovalcev zelo redke in morda predstavljajo okoli 5 - 10 % vseh poizvedb.

Vse poizvedbe, ki označujejo določene človeške lastnosti ali logične postopke/ravnanja, se prav tako razvrščajo pod UDK 1. Poizvedba kot je npr. "Etika in morala", se razvršča kot UDK 1, toda v primeru, da se poizvedba glasi "Etika in morala družbe", se tovrstna poizvedba razvrsti pod UDK 3, kajti pojma kot sta "Etika in Morala" postaneta podrejena pojmu "Družba". Tovrstni ipd. primeri veljajo kot IZJEME!

3. UDK 3 - npr. poizvedba "Družba v stari Grčiji" razvrščam pod UDK 3, a poizvedbo "Stara Grčija" pod UDK 9 (*velja prednostno pravilo dogovorjenega razvrščanja po UDK glavnih tablic*). Ta primer s prakse je bil nekoliko drugačen, kajti spletni obiskovalec je najprej izvedel poizvedbo "Družba v stari Grčiji" in nato šele poizvedbo "Stara Grčija", vendar je prvo poizvedbo izvedel dvakrat, medtem ko drugo zgolj enkrat. Spletni obiskovalec je obiskal področje UDK 3 in področje UDK 9. Kako ta primer razrešiti? Razrešimo ga tako, da ugotovimo namen spletnega obiskovalca in povrh tega upoštevamo utež tj. število posamezne poizvedbe. Na podlagi razmerja 2 poizvedbi proti 1 poizvedbi (2:1) sem obe poizvedbi razvrstil pod UDK 3.

5. UDK 4 - npr. poizvedba "... in zman" bi lahko pomenila del besedila določenega spletnega dokumenta, vendar tega kot klasifikator ne vemo, zato razvrstimo tovrstno poizvedbo pod UDK 4. Poizvedba "Razpi" je nepopolna in bi lahko pomenila okrajšavo za pojem "Razpis" ali "Razpisi", vendar se kot klasifikator ne moremo dokopati do pravega namena oziroma pomena tovrstne poizvedbe in zato se razvrščajo takšne poizvedbe kot UDK 4. Prav tako se v to UDK 4 razvrščajo poizvedbe, kjer gre za nepoznane in nerazrešljive pojme kot npr. »Avrikeoj«. V zelo redkih primerih gre celo za takšne poizvedbe v tujem jeziku, ki ga kot klasifikator ne prepoznamo in zaradi tega takšne poizvedbe prav tako razvrstimo pod UDK 4.

6. UDK 5 - npr. poizvedba "Naravni antibiotik" gre pod UDK 5, vendar z vidika praktične uporabe naravnega antibiotika lahko tovrstno poizvedbo razvrstim pod UDK 6, kajti naravni antibiotiki se uporabljajo za zdravljenje oziroma uničevanje nezaželenih bakterij (*zdravstvo* -

UDK). Pri tovrstnih poizvedbah ipd. moram nujno preveriti še ostala kazalca (*podobno kot pri primeru gl. 2.a.*).

7. UDK 6 - v to področje razvrstim poizvedbe spletnih obiskovalcev, s katerih je nazorno razvidno, da gre za bolezni, zdravljenja, zdravilne pripomočke, anatomijo človeka, zdravstvene metode, stroje, električna omrežja, gradbena dela, industrije itd. Npr. poizvedba "Elektrika" se razvršča kot UDK 5 (*teoretični vidik elektrike*), medtem ko se poizvedba "Elektrotehnika" razvrsti pod UDK 6 (*uporabni ali praktični vidik elektrike*).

8. UDK 7 - npr. poizvedba "Kultura" je splošne narave in jo s tega vidika razvrščam pod UDK 0, toda ne v primerih, ko se poizvedba glasi "Kultura umetnosti", "Kultura v športu", "Igralna kultura" ipd., kajti te poizvedbe vedno razvrščam pod UDK 7. Npr. poizvedbo "Arhitektura Šola Izobraževanje" (*trobasedna poizvedba*) se razvršča kot UDK 3 in ne kot UDK 7, ker gre v tem primeru za izobraževalno dejavnost, ki je v povezavi z arhitekturo.

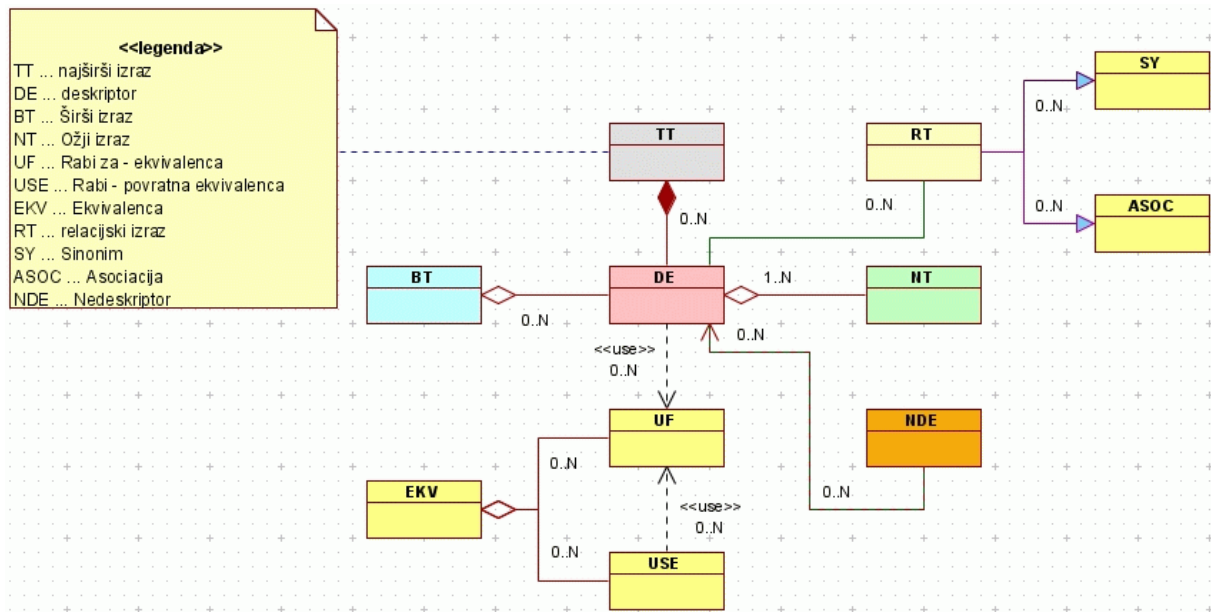
9. UDK 8 - Npr. poizvedba "Besedna komunikacija" se razvršča kot UDK 8, ker gre v tem primeru za težišče k jeziku, toda poizvedbo "Komunikacija" razvrstim pod UDK 0 (*je splošne narave, ne opredeljuje posebnost*), medtem ko pa poizvedbo "Socialna komunikacija" razvrščam pod UDK 3 (*izpostavljena je posebnost "Socialna"*).

10. UDK 9 - V primeru poizvedbe "Reka" ali "Jezero" razvrstim obe poizvedbi pod UDK 5, toda ne v primerih, ko gre za naslednji poizvedbi: "Reka v Braziliji", "Jezero v Sloveniji". Tovrstne poizvedbe vedno razvrščam pod UDK 9, ker so takšne ipd. poizvedbe še krajevno opredeljene!

Po dosedanjih izkušnjah na področju UDK klasificiranja poizvedb spletnih obiskovalcev lahko trdim, da je okoli 90 % odstotkov poizvedb nedvoumne narave in okoli 5 do 10 % odstotkov poizvedb je takšnega značaja, da je potrebno poleg UDK glavnih tablic in geselskega kataloga uporabiti še druge pripomočke in/ali druge kazalce. Prav tako so zelo redke poizvedbe, ki jih lahko razvrstimo zgolj pod UDK 4, in celo nekatere od teh je možno v poznejšem obdobju razrešiti (*npr. ugotovimo pomen nepoznane besede*). Sledi kratek opis tezavra, in možnih pojmovnih odnosov v njemu, nakar bom še omenil vrste tezavrov.

## 2.2.5 Tezaver

Tezaver je standardizirani referenčni slovar, ki je zbirka besed naravnega jezika (*splošnega in strokovnega*) s predstavitvijo njihovih pojmovnih odnosov in nam služi kot pomoč pri dokumentaciji. Tezaver lahko ponazori hierarhične, ekvivalentne in asociativne odnose med besedami. Tezaver je lahko tudi v vizualni ali slikovni obliki. Naj v tem delu še predstavim slikovni prikaz, ki ponazarja različne odnose med pojmi.



### 2.2.5.1 Slikovni prikaz 2: Možna različica odnosov med pojmi v tezavru

Slika 2 prikazuje možno različico odnosov med pojmi v tezavru in legendo. V sredini je deskriptor (v nadaljevanju *DE*), ki je do drugih pojmov v različnih odnosih. Odnosi so lahko hierarhični (*TT*, *BT*, *NT*), ekvivalentni (*UF*, *USE*) in asociativni (*RT*, *SY*, *ASOC*):

- *TT* ... najširši ali supernadrejen izraz (v primeru, da je *DE* enak *TT*, potem *DE* ne vsebuje *BT* ali širšega izraza)
- *BT* ... širši ali nadrejen izraz v odnosu do *DE*
- *NT* ... ožji ali podrejen izraz v odnosu do *DE*
- *UF* ... rabi za ali izražen ekvivalenten odnos *DE*, ki je najznačilnejši do drugega *DE*
- *USE* ... rabi ali izražen ekvivalenten odnos, ki je manj pogost oziroma značilen do drugega *DE*
- *RT* ... relacijska povezava do *DE* (v tem primeru sem *RT* generaliziral še na *SY* ali sinonimni izraz in na asociativni izraz), skratka *RT* = *SY* = *ASOC*
- Na koncu obrazložitve slikovnega prikaza 2, je potrebno še opozoriti na razred *NDE* ali nedeskriptor, ki je pomensko gledano enačajejen z *DE*, vendar se *NDE* manj pogosto uporablja ali pa je manj strokoven in bolj poljudne ali pogovorne narave.

Odnosi med pojmi so lahko še mnogo bolj kompleksni, še zlasti z vidika vrtnja po hierarhiji navzdol, kar pa naj ne bi bilo iz informativnega dokumentalističnega vidika razen v redkih primerih zelo smiselno.<sup>17</sup> Iz metodološkega vidika tezavre lahko členimo na analitične ali razčlenjevalne in na sintetične ali združevalne. Z ozirom na namen uporabe tezavra poznamo mikrotezavre, makrotezavre, mezotezavre, fasetne in multidimenzionalne tezavre idr. Izraza fasetni in multidimenzionalni sta si nekoliko sorodna, vendar se kljub temu razlikujeta, kajti multidimenzionalni tezavri v odnose vključujejo več različnih dimenzij, ki so si med seboj lahko povsem enakovredni in v odnosu do DE nadrejeni, medtem ko fasete zgolj dopolnjujejo odnose med DE in drugimi pojmi. Sicer se pojem multidimenzionalni tezaver zelo redko uporablja. O multidimenzionalnem tezavru bom natančneje poročal v enem od zadnjih poglavjih tega doktorskega dela, ko bom predstavil kvalitativni model adaptivne digitalne knjižnice.

V nadaljevanju bom opisal področje podatkovnega rudarjenja (*angl.: Data Mining*) oziroma odkrivanja zakonitosti v podatkih in ob tem še zlasti omenil tiste metode/tehnike, ki jih nameravam v tem doktorskem delu uporabiti.

## 2.3 Odkrivanje zakonitosti v podatkih

Začetki odkrivanja zakonitosti v podatkih (*v nadaljevanju: OZVP*) so v bistvu že zaznavni pri klasični statistiki, vendar je sam pojem šele nastal v zgodnjih 90- tih letih prejšnjega stoletja.

Razvoj OZVP je tako šel skozi tri stopnje:

1. Klasična statistika – regresijske analize, diskriminantne analize, standardna distribucija, standardni odklon, variance, analize grozda/clustrov idr. Klasična statistika pomeni osnova za OZVP.
2. Umetna inteligenca – heuristične analize, posnemanje človekovega mišljenja, reševanje problemov, izgradnja ustreznih izvedb ipd.
3. Strojno učenje – združevanje spoznanj in metod s področja klasične statistike in umetne inteligence, računalniški programi se učijo iz podatkov, razvijanje kompleksnih odločitvenih modelov in programskih izvedb v poslovne namene idr.<sup>18</sup>

"OZVP je področje, ki temelji na metodah in algoritmih umetne inteligence in njenega ožjega področja strojnega učenja." OZVP poskuša tako odkriti značilne vzorce, pravila in trende v

---

<sup>17</sup> Tovrstno priporočilo sem prebral v dveh monografskih publikacijah: Urbanija, J.(1996). Metodologija izdelave tezavra in Broughton, V.(2006). Essential Thesaurus construction.

<sup>18</sup> Izpeljava: Bratko, I., Kubat, M. & Ryszard, S.(1998). MACHINE learning and data mining : methods and applications. Chichester [etc.] : J. Wiley & Sons, XVI, 456 str.

podatkih z velikim obsegom, da bi potem lahko iz njih izluščili določena spoznanja in v naslednji stopnji nova znanja.<sup>19</sup> Tovrstna na novo odkrita oziroma izpeljana znanja se potem lahko uporabijo za izdelavo različnih vrst izvedb od programskih orodij, spletnih trgovin pa tja do smotrno urejenih digitalnih knjižnic (*informacijski sistemi za upravljanje z nanjem - organizirane informacije, ki so prilagojene potrebam uporabnikov in/ali strankam za različne pomembne odločitve*) idr. Podpodročji odkrivanja zakonitosti v besedilih (*angl. Text Mining*) kot tudi odkrivanje zakonitosti v numeričnih podatkih spadata prav tako v področje OZVP. Pozneje bom še nekoliko predstavil področje odkrivanja zakonitosti v besedilih in omenil programsko orodje AntConc. S pomočjo tega programskega orodja bom pripravil podatke za uporabo nekaterih metod in vizualizacijskih tehnik s področja OZVP. Za uporabo teh metod in vizualizacijskih tehnik bom uporabil svetovno znano programsko orodje Orange Canvas. Z vse močnejšo uveljavitvijo interneta oziroma svetovnega spleta je nastala močna podveja OZVP, ki so jo poimenovali kot spletno (*podatkovno*) rudarjenje (*angl.: Web mining*), ki se členi na spletno vsebinsko rudarjenje (*angl.: Web Content Mining*), spletno strukturno rudarjenje (*angl.: Web Structure Mining*) in spletno uporabniško rudarjenje (*angl.: Web Usage Mining*). Spletno uporabniško rudarjenje se ukvarja z odkrivanjem zakonitosti v podatkih na spletu, ki so nastali kot izid različnih dejavnosti (*npr. izdelava profilov uporabnikov, uporaba iskalnika oziroma poizvedovanje, kliki na spletne povezave, prenos datotek s spleta na disk*) spletnih obiskovalcev po spletnih straneh.

### **2.3.1 Nekatere metode in tehnike podatkovnega rudarjenja ali odkrivanja zakonitostih v podatkih**

S splošnega vidika, se metode OZVP delijo na opisne in napovedovalne.

Opisal bom zgolj tiste metode, ki jih bom pri kasnejši obdelavi podatkov tudi uporabil. Uporabil bom naslednji metodi in vizualizacijske tehnike OZVP: asociacijska pravila in razvrščanje v skupine (*metode*); distribucija, prikaz statističnih atributov, Scatter Plot in matriko, Sievov diagram, dendrogram in distančno mapo (*vizualizacijske tehnike*).

### **2.3.2 Klasifikacija**

Metoda klasifikacije pri OZVP je tako opisne kot tudi napovedovalne narave. Sestavljena je iz dveh stopenj. V prvi stopnji se opredelijo lastnosti podatkov, kar se lahko prikazuje z

---

<sup>19</sup> Opredelitev sem prevzel na podlagi naslednjega članka: Rupnik, R. & Krisper, M.(2004). Aplikativni sistemi za odkrivanje zakonitosti v podatkih na str. 2.

odločitvenimi drevesi ali pravili. V drugi stopnji se predhodno pridobljeni model uporablja za razvrščanje novih primerov v razrede. V tem delu bom v spremnem gradivu na CD-ROM podal primerjavo med klasifikacijsko metodo Naive Bayes in klasifikacijskim drevesom. V papirnem gradivu pa o metodi klasifikacije pri OZVP ne bom natančneje poročal, ker menim, da mi bosta dajali (z ozirom na pripravljene podatke, vsebinskega poudarka idr.) metodi asociacijska pravila in še zlasti razvrščanja v skupine (*angl.: Clustering*) kakovostnejše oziroma nazornejše izide.

### 2.3.3 Asociacijska pravila

Metoda asociacijska pravila (*Association Rules*) je opisne narave, ki omogočajo prepoznavanje razmerij, povezav, asociacije med vrednostmi oziroma lastnostmi znotraj preučevane podatkovne množice. Najbolj znan primer uporabe asociacijskih pravil predstavlja analiza nakupovalne košarice. Pri tej analizi lastniki spletnih trgovin npr. na podlagi nakupov strank ugotavljajo, katere informacije o izdelkih na spletni strani je smiselno ponuditi skupaj, da bi s tem povečali prodajo. Nekateri izdelki so z uporabnega vidika stranke tako zelo povezani med sabo, da je potrebno ponudbo izdelkov organizirati tako, da se bodo stranke takšnih povezanosti zavedale in na takšen način kupile ne zgolj en izdelek, temveč še druge izdelke (*npr. stranka kupi čips in poleg tega nabavi tudi pivo*). Takšno asociacijsko pravilo potem lahko splošno zapišemo na naslednji način:

$$X \rightarrow Y \text{ (zanesljivost, podpora)}$$

Določenejšo lahko to pravilo zapišemo na naslednji način:

$$\text{Čips} \rightarrow \text{Pivo} (0,80 \ 0,07)$$

Ta kratki zapis ima naslednji pomen: v primeru, da je stranka kupila čips, je v 80,00 % primerov kupila tudi pivo. Delež strank, ki so kupile oba izdelka, se je pojavilo v 7,00 % vseh prenosov ali transakcij. V tem delu se bom ukvarjal z naslednjim vprašanjem, ki je tesneje povezano z upravljanjem znanja oziroma z ustrezno organizacijo podatkov/informacij, tako da lahko spletni obiskovalci lažje najdejo več koristnih stvari na enem mestu in pri tem ne izvedejo nepotrebne akcije, tako da ne izgubijo po nepotrebem čas in energijo: "Katere podatke/informacije in/ali storitve je smiselno ponuditi skupaj oziroma fizično lokacijsko gledano eno blizu drugo?"<sup>20</sup> Ugotavljanje asociacijskih pravil glede večje ali manjše povezanosti določenih podatkov/informacij s strani spletnih obiskovalcev pomeni lahko tudi

---

<sup>20</sup> Zamisel sem izpeljal na podlagi naslednjega članka: Rupnik, R. & Krisper, M.(2004). Aplikativni sistemi za odkrivanje zakonitosti v podatkih na str. 3.

nekakšen dodaten pripomoček za razvrščanje spletnih obiskovalcev v posamezne skupine ali grozde in celo moč povezanosti med različnimi grozdi le-teh.

### **2.3.4 Razvrščanje v skupine**

Razvrščanje v skupine pri OZVP je prav tako opisne narave. Pri tej metodi želimo ugotoviti značilne predstavnike znotraj velike množice podatkov. Različne skupine znotraj velike množice se med seboj razlikujejo po lastnostih, medtem ko so si predstavniki znotraj skupine po lastnostih podobni, kljub dejstvu, da obstajajo znotraj teh manjših množic hierarhije. Na podlagi te metode je možno ugotoviti s katerimi skupinami spletnih obiskovalcev imamo opraviti in povrh tega lahko posredno dokaj podrobno odkrijemo njihove vzgibe in/ali potrebe. Skratka ta metoda nam lahko posreduje vedenje o miselnih modelih/hierarhijah manjših in večjih skupin spletnih obiskovalcev. V nadaljevanju bi bilo možno primerjati značilnosti tovrstnih skupin z značilnostmi spletnih obiskovalcev posameznikov (*npr. dejavnosti posameznih spletnih obiskovalcev na določeni izobraževalni zbirki znanja ipd., kar pa že spada v področje spletnega uporabniškega rudarjenja*). V tem delu se bom predvsem ukvarjal z razčlenbo interesnih skupin spletnih obiskovalcev.

### **2.3.5 Vizualizacijske tehnike**

Uporabil bom naslednje že omenjene vizualizacijske tehnike za prikaz podatkov: distribucija, prikaz statističnih atributov, Scatter Plot, Sievov diagram, dendrogram in prikaz distančne mape. Pri tem bom uporabil programsko orodje Orange Canvas.

#### **2.3.5.1 Distribucija**

S pomočjo slikovne tehnike distribucije je možno ponazoriti frekvenčno porazdelitev množice podatkov. Povrh tega lahko dodatno pod "Settings (*nastavitve*)" prikažemo krivuljo verjetnosti (*angl.: probability – verjetnost*) in zaupanja (*angl.: confidence interval - pomeni sorazmerna vrednost množice, ki je zanesljiva oziroma, kateri je možno še zaupati*).

#### **2.3.5.2 Statistični atributi**

Z opcijo statistični atributi je dana možnost, da slikovno in v številkah prikažemo lastnosti podatkovne množice (*povprečje, mediano, minimalne vrednosti, maksimalne vrednosti, posamezne frekvence*). S pomočjo te možnosti si bom prihranil veliko časa pri pripravi podatkov za prikaz zakona moči.

### 2.3.5.3 Scatter Plot

S pomočjo Scatter Plot-a je možno slikovno med seboj primerjati različne lastnosti in razrede in s tem prikazati različne odvisnosti oziroma razmerja med njimi (*hiperbolična odvisnost med spremenljivkama, parabolična odvisnost med spremenljivkama, linearne odvisnosti idr.*).

### 2.3.5.4 Sievov diagram

Sievov diagram ponazarja frekvence v dvosmerni kontingenčni tabeli z ozirom na relacijo neodvisnih pričakovanih frekvenc, nakar izpostavlja asociacijske vzorce med stolpčnimi in vrstičnimi spremenljivkami. V tem delu bom Sievov diagram uporabil za prikaz gostosti poizvedb po posameznih področjih človekovega znanja.

### 2.3.5.5 Dendrogram<sup>21</sup>

Podatkovno hierarhično zgradbo grozda (*ali clustra*) je možno slikovno ponazoriti s pomočjo dendrograma. Dendrogram (*grško: Dendros = drevo*) v bistvo posnema zgradbo drevesa, kjer se hierarhično razgradijo podatkovne množice v čedalje manjše enote. Korenina drevesa pomeni temeljni grozd, listja drevesa predstavljajo posamezne skupine, medtem ko vozlišča predstavljajo združitev določenih podskupin. V tem vpogledu je pomembna funkcija oddaljenosti tako med posameznimi objekti kot tudi med množicami oziroma količinskimi enotami različnih objektov. S pomočjo dendrograma bom ponazoril interesna področja spletnih obiskovalcev in s tem posredno tudi povezanost med različnimi področji človekovega znanja (*segmentacija ali razčlenitev spletnih obiskovalcev po področjih zanimanja*).

### 2.3.5.6 Distančna mapa

V tem delu bom slikovni prikaz distančne mape uporabil za ponazoritev tako oddaljenosti kot tudi povezanosti oziroma bližine med posameznimi UDK področij, ki so izid poizvedb spletnih obiskovalcev na iskalniku UDK leksikona, OZVP besednega dela in OZVP numeričnega dela poizvedb. Prav na podlagi uporabljenih kazalcev kot so frekvenca (f), rang (r) in moči ključne besede (K), sem lahko na lestvici od nič do pet ovrednotil moč ali šibkost povezanosti med različnimi UDK področij. Prav tako bom tudi o tem postopku natančneje poročal v naslednjem poglavju.

## 2.3.6 Odkrivanje zakonitosti podatkov v besedilih<sup>22</sup>

Odkrivanje zakonitosti podatkov v besedilih (*angl.: Text Mining*) pomeni raziskovanje novih, pravih, relevantnih podatkov/informacij v besedilih, ki je še zlasti v današnjem času

---

<sup>21</sup> Podatke sem izpeljal na podlagi naslednjega vira: <http://de.wikipedia.org/wiki/Dendrogramm> (2005-07-28)

<sup>22</sup> Podatke sem izpeljal na podlagi naslednjega vira: <http://de.wikipedia.org/wiki/Textmining> (2005-05-12)

podprto z učinkovitimi in kakovostnimi računalniškimi programskimi orodij. Bistvena razlika med OZVP v besedilih in OZVP je v tem, da so podatki pri OZVP v besedilih manj strukturirani kot pri OZVP. Pri OZVP v besedilih velja težnja, da je potrebno nestrukturirane ali manj strukturirane podatke pripraviti tako, da jih preoblikujemo v bolj strukturirano obliko, nakar lahko potem s postopkom/-i OZVP tovrstne podatke nadalje obdelujemo. Jezikoslovne analize navadno potekajo tako, da je potrebno celotno besedilo razgraditi na manjše gradnike kot npr. na stavke, besedne zveze (*fraze*) in posamezne besede. V tem delu bom uporabil določeno metodo OZVP v besedilu tako, da bom kot besedni korpus podal zbrane poizvedbe, ki so jih spletni obiskovalci izvedli v obdobju treh let, nakar bom te podatke s pomočjo programskega orodja za besedno analizo AntConc obdelal z naslednjimi statističnimi kazalci: rang, frekvenca in dejavnik pomembnosti ključnih besed (*angl.: Keyness*). Še zlasti z dejavnikom pomembnosti ali moči določene ključne besede nameravam odstraniti vse tiste stavke, besedne zveze in besede, ki se v korpusu poizvedb in kontekstov redkeje in po pomenski strani manj izrazito pojavljajo. Skratka korpus poizvedb bom pripravil tako, da bom te sorazmerno šibko strukturirane podatke pretvoril v strukturirane. Prav s tem bom lahko v nadaljevanju svojih prizadevanj tovrstne podatke obdelal s pomočjo določenih metod OZVP. Ta postopek nameravam v naslednjem poglavju (*tj. pri pripravi podatkov*) natančneje obrazložiti. OZVP v besedilih se lahko uporablja na različnih področjih tako za poslovne (*npr. za izdelavo iskalnika v finančnem sektorju idr.*) kot tudi za obogatitev izobraževalnega procesa (*npr. kot didaktični pripomoček za lažje usvajanje učne/študijske snovi idr.*). Na priloženem CD-ROM bom še kot zanimivost posredoval izvedeno analizo gostoto besed in besednih zvez v korpus zbranih poizvedb. V naslednjem podpoglavju, bi želel nekoliko izpostaviti pomen statističnega kazalca "Pomembnost ali moč ključne besede" (*K – Keyness*).

### **2.3.6.1 Pomembnost ali moč določene ključne besede<sup>23</sup>**

Pomembnost ali moč ključne besede (*v nadaljevanju K*) nam posreduje povratno informacijo o tem, katere besede dejansko pomembno označujejo določen besedni korpus. Pri izračunu **K** določene ključne besede v besedilu, se upoštevajo naslednji kazalci:

- frekvenca ključne besede v manjšem seznamu besed
- vsota vseh besed v manjšem seznamu besed
- frekvenca ključne besede v referenčnem korpusu besed
- vsota vseh besed v referenčnem korpusu

---

<sup>23</sup> [http://filolog.uni.lodz.pl/elise/emodule/12\\_keyness.html](http://filolog.uni.lodz.pl/elise/emodule/12_keyness.html) (2006-01-10)

S programskim orodjem za besedne analize (*oziroma OZVP v besedilih*) AntConc je možno ključnim besedam določiti **K**, vendar je potrebno predhodno podatke ustrezno pripraviti, kar bom prav tako predstavil, še zlasti v podpoglavju 3.8.1 na strani 58.

## 2.4 Digitalne knjižnice

V tem podpoglavju ne nameravam poročati o različnih opredelitvah in praktičnih primerih digitalnih knjižnic (*v nadaljevanju DK*), ampak bo to podpoglavje predvsem posvečeno zamisli o dodatnih aplikativnih modulih, ki naj bi pomagali oziroma olajšali uporabniku izbor pomembnih informacij. Pri sprotni opombi 24 si bo lahko bralec tega dela ogledal naslov vira, v katerem so podrobneje opisane različne opredelitve in vidike DK. Pri sprotnih opombah od 25 do 27 se bo lahko bralec seznanil z viri, ki vsebujejo različne modele in aplikativne pristope, ki bodo verjetno tudi v prihodnje za DK pomembne.

DK so informacijski ravnateljevalni sistemi (*Information Management Systems*), ki oskrbujejo uporabnike s pomembnimi informacijami in/ali znanjem z namenom, da olajšajo nekatere odločitve. Podatki na DK naj bi bili tako organizirani, da oskrbujejo z informacijami tako zunanje kot tudi poslovne ali notranje uporabnike z namenom, da se čim bolj gospodarno in učinkovito odvija delovni/poslovni proces izmenjave informacij v določenem podjetju ali zavodu. Uporabniki naj bi prišli do pomembnih informacij za odločitve po čim krajši poti in s manjšo porabo energije. DK v bistvu že ponujajo določena pripravljena/organizirana znanja, ki naj bi bila prilagojena profesionalnim in drugim uporabnikom.

V preteklosti so bile DK pretežno zelo ohlapno opredeljene kot nekakšne amorfne discipline, ki jih je možno vključiti kot nekakšne dodatne informativne pripomočke za uporabnike z različnih interesnih področij.<sup>24</sup> Glede prihodnostnega razvoja DK je potrebno izpostaviti dejstvo, da DK ne morejo biti zgolj v domeni bibliotekarske znanosti in stroke kot je bilo temu bolj ali manj primer do sedaj, ampak se bo povečal vpliv drugih znanosti in tehničnih ved, še zlasti računalništva in informatike. DK prihodnosti bodo po vsej verjetnosti še bolj vključevale različne aplikativne module, ki bodo omogočali tako socialno zблиževanje, profesionalno komunikacijo med ljudmi kot tudi raziskovanje informacij in uporabnikov.<sup>25</sup> Z mojega vidika pomeni torej DK nekakšen informacijski sistem, ki je sestavljen iz različnih aplikativnih modulov (*npr. notranji inteligentni iskalnik, aplikacije za spremljanje dejavnosti*

---

<sup>24</sup> Nekoliko več o opredelitvah digitalnih knjižnic si lahko preberete na naslednji spletni strani, ki predstavlja del mojega e- magistrskega dela: <http://udkclick.bravehost.com/mag3.htm>.

<sup>25</sup> Lynch, C.(2005). Where do we go from here? : the next decade for digital libraries. V: D-Lib Magazine, Vol. 11, No. 7/8. Članek je v elektronski obliki in je dosegljiv na naslednjem URL: <http://www.dlib.org/dlib/july05/lynch/07lynch.html> (2006-12-29).

spletnih obiskovalcev, diskusijski forumi, tezavri itd.) za sprejemanje različnih odločitev, kar se še zlasti izvaja na podlagi preučevanja uporabnikov določene DK, z namenom, da se izdelajo (*interesni*) profili le-teh. Prav na podlagi izdelanih informacijskih/interesnih profilov uporabnikov je možno izdelati takšne DK, ki se prilagaja in učinkovito podpira različne vrste uporabnikov.<sup>26</sup> Prav na podlagi poprej zapisanemu, je možno izpeljati dodatno dejstvo, da se lahko DK uporabljajo v različnih delovnih/poslovnih okolij (*npr. vladne institucije, znanstvene raziskovalne institucije, vzgojnoizobraževalni zavodi, industrija itd.*).<sup>27</sup> Na DK se bom še v poglavju 5 povrnil, ko bom prikazal končni model adaptivne DK, ki si jo zamišljam kot aplikativni sistem, kateri zmore kakovostno ter učinkovito podpreti vzgojnoizobraževalne in znanstveno raziskovalne procese in s tem posledično tudi odgovarjajoče uporabnike. V podpoglavju »Možnosti uporabe izpeljanih novih znanj/spoznanj« nameravam poročati o aplikativnih modulih DK; še zlasti o aplikativnem modulu multidimenzionalnega tezavra (*moduli bodo prikazani znotraj razvitega modela adaptivne DK*), ki naj bi jih adaptivna DK vsebovala.

## **2.5 Izdelava interesnih profilov uporabnikov na podlagi preučevanih dejavnosti uporabnikov na pilotni digitalni knjižnici**

Na pilotni digitalni knjižnici sem spremljal oziroma meril vrsto poizvedb in njih frekvenco, ki so jih izvedli spletni obiskovalci na notranjem iskalniku ter število obiskov po različnih področnih spletnih strani. S pomočjo webometrije in nekaterih metod s področja OZVP bom v tem delu izdelal interesne profile uporabnikov. Z ozirom na kasnejše preučevanje poizvedb, bo potrebno predhodno še odgovoriti na naslednje vprašanje »Zakaj nekateri uporabniki sploh uporabljajo iskalnike?« Prav odgovor na to vprašanje bo dodatno osvetlil določene informacijske vedenjske vzorce uporabnikov, kar mi bo v nadaljnji stopnji olajšalo izdelavo interesnih profilov uporabnikov na pilotni DK.

### **2.5.1 Spletni iskalniki**

V vsej svoji evoluciji se je človek posluževal različnih orodij, da se je lahko dokopal do zaželenega predmeta in/ali zamisli ter tako s tem razrešil bolj ali manj močne življenjske probleme. Prav ta nespreminjajoči vsesplošni vedenjski vzorec človeka se tudi do danes ni

---

<sup>26</sup> Coleman, A. & Summer, A. (2004). Digital libraries and user needs : negotiation the future. *Journal of Digital Information*, Volume 5 Issue 3. URL: <http://jodi.tamu.edu/Articles/v05/i03/editorial> (2006-12-29).

<sup>27</sup> Na naslednji spletni strani si je možno ogledati različne modele DK: <http://www.scils.rutgers.edu/~tefko/ProcASIST2001.doc> (2006-12-30)

spremenil, čeprav se je v mnogih primerih spremenila preobleka oziroma oblika določenih orodij in akcij, ki jih uporablja ter izvaja današnji sodobni človek v sodobnih človeških družbah. V današnjem času svetovnega spleta lahko poročamo o določeni skupini ljudi, ki jih označujemo kot spletne obiskovalce. Prav ti lahko imajo še vrsto drugih oznak, vendar je vsem skupno to, da uporabljajo svetovni splet kot orodje tako za krajšanje prostega časa, ustvarjanje izdelkov, za komunikacijo idr. in še zlasti za razreševanje bolj ali manj nujnih življenjskih problemov. Nesmiselno bi bilo zdaj naštet štampilna orodja, ki jih je možno na svetovnem spletu uporabiti, tako da bom na tej točki izpostavil zgolj eno, ki je tudi eno od pomembnih zanimanj tega doktorskega dela, tj. spletni iskalnik. Spletni iskalnik je orodje, s katerim si je možno na svetovnem spletu poiskati zaželene podatke in/ali informacije. Ob tej ugotovitvi tudi ni več daleč do ključnega vprašanja o tem, kdaj oziroma v katerih situacijah spletni obiskovalci sploh uporabljajo omenjeno orodje? Prav s tega temeljnega vprašanja lahko izpeljemo številna druga, ki so npr. naslednja:

- Kakšen je ta temeljni miselni in akcijski vzgib oziroma svet te skupine ljudi?
- Na kakšen način oziroma katere tehnike uporabljajo spletni obiskovalci, da se dokopljejo do potrebnih podatkov/informacij?
- Kako so tovrstni ljudje zadovoljni z organizacijo podatkov/informacij v določeni digitalni oziroma spletni zbirki znanja?
- Kakšno je razmerje med ponudbo in povpraševanjem podatkov/informacij na zbirkah znanj ter njihovimi informacijskimi potrebami?
- S kakšnimi znanji in tehnologijami razpolagajo tovrstni ljudje?
- Kakšna je globalna miselna hierarhija spletnih obiskovalcev z ozirom na njihova področja zanimanja?
- V kolikšni meri se organizirane informacije na različnih zbirkah znanj prilegajo potrebam te skupine ljudi in v kolikšni meri vplivajo t.i. pripravljene pametni podatki na izide izvedenih poizvedb s pomočjo določenega spletnega iskalnika?
- Kakšne so vizije oziroma prihodnostne silnice spletnih obiskovalcev glede doseganja lastnih ciljev?
- Kakšne so predstave spletnih obiskovalcev o spletnih iskalnikih oziroma kakšni naj bi bili?
- Kakšna je interpretacija oziroma obdelava podatkov/informacij spletnih obiskovalcev?
- ...

Ob vseh teh vprašanjih idr. je možno izpeljati paradoks, ki poroča o bogati ponudbi podatkov/informacij kot tudi o težavnosti orientiranja spletnih obiskovalcev, da najdejo zaželeno, kajti večja kot je podatkovna/informacijska ponudba na določeni spletni zbirki

znanja, toliko zapletenejše bodo tudi akcije oziroma navigacije spletnih obiskovalcev po določenih spletnih straneh.<sup>28</sup> Slednje navedeno je močen argument za spletne obiskovalce, da se bodo posluževali orodja, ki jim bo olajšalo orientacijo ali navigacijo po spletnih straneh z bogato ponudbo podatkov/informacij v soodvisnosti s povpraševanjem z njihove strani.

Na tovrstna razmišljanja, se bom še povrnil v poznejših poglavjih.

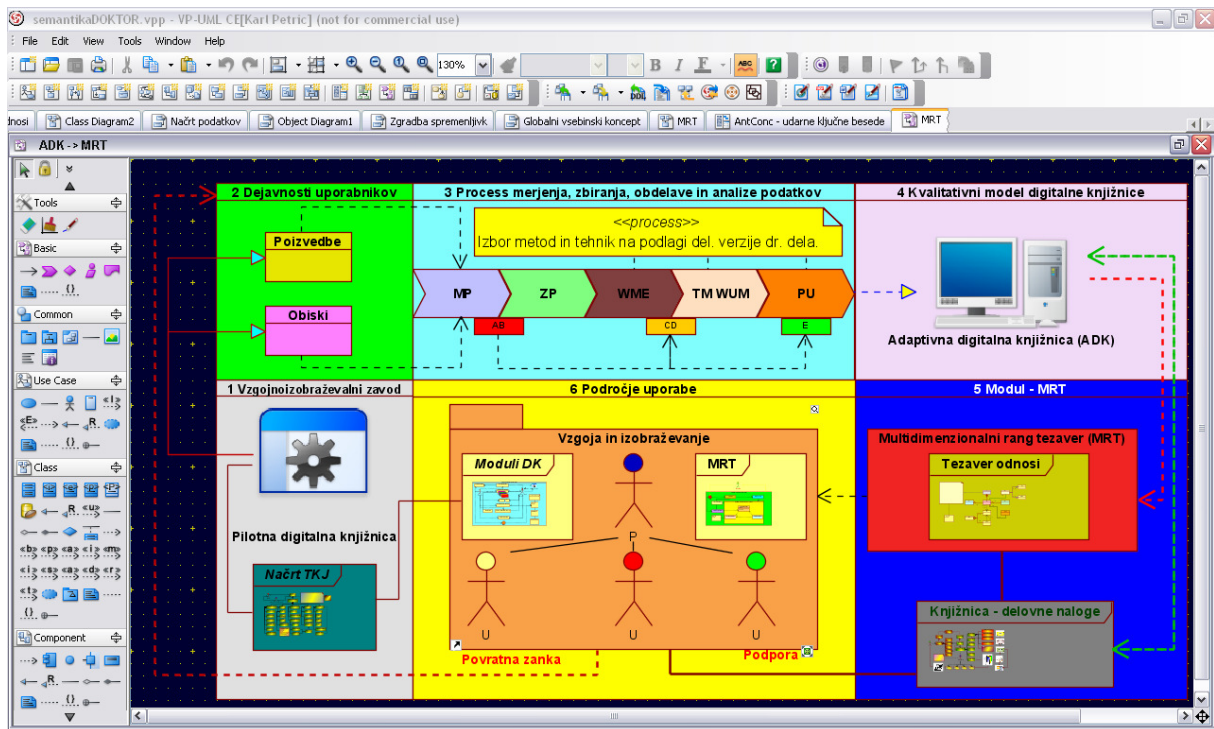
V nadaljevanju tega dela nameravam predstaviti snovanje oziroma zasnovano kvalitativnega modela adaptivne digitalne knjižnice s pomembnim modulom multidimenzionalnega rang tezavra, nakar bom znotraj zasnovanega načrtoval in pripravil podatke za kasnejši OZVP.

---

<sup>28</sup> Misel sem izpeljal na podlagi naslednjega spletnega dela: <http://www.searchtools.com/guide/index.html> (2006-12-2005).

# 3 Analiza za pripravo modela

## 3.1 Zasnova modela



### 3.1.1 Slikovni prikaz 3: Snovanje kvalitativnega modela modela adaptivne digitalne knjižnice z multidimenzionalnim rang tezavrom

Slika 3 v strnjeni obliki ponazarja vse postopke in procese, ki sem jih izvedel, da sem lahko zasnoval kvalitativni model ADK, ki izpostavlja pomemben modul multidimenzionalnega rang tezavra (*MRT*) za podporo poslovnih in drugih uporabnikov v vzgojnoizobraževalnem procesu. Posamični mejniki (*gl. slika 3*) so oštevilčeni od 1 do 6. Naj na kratko povzamem bistvo tovrstnih mejnikov:

Mejnik 1 – sociološko proučevanje srednješolskega zavoda z vidka bibliotekarske stroke in prakse, izdelava načrta, nakar vzpostavitev pilotne digitalne knjižnice.

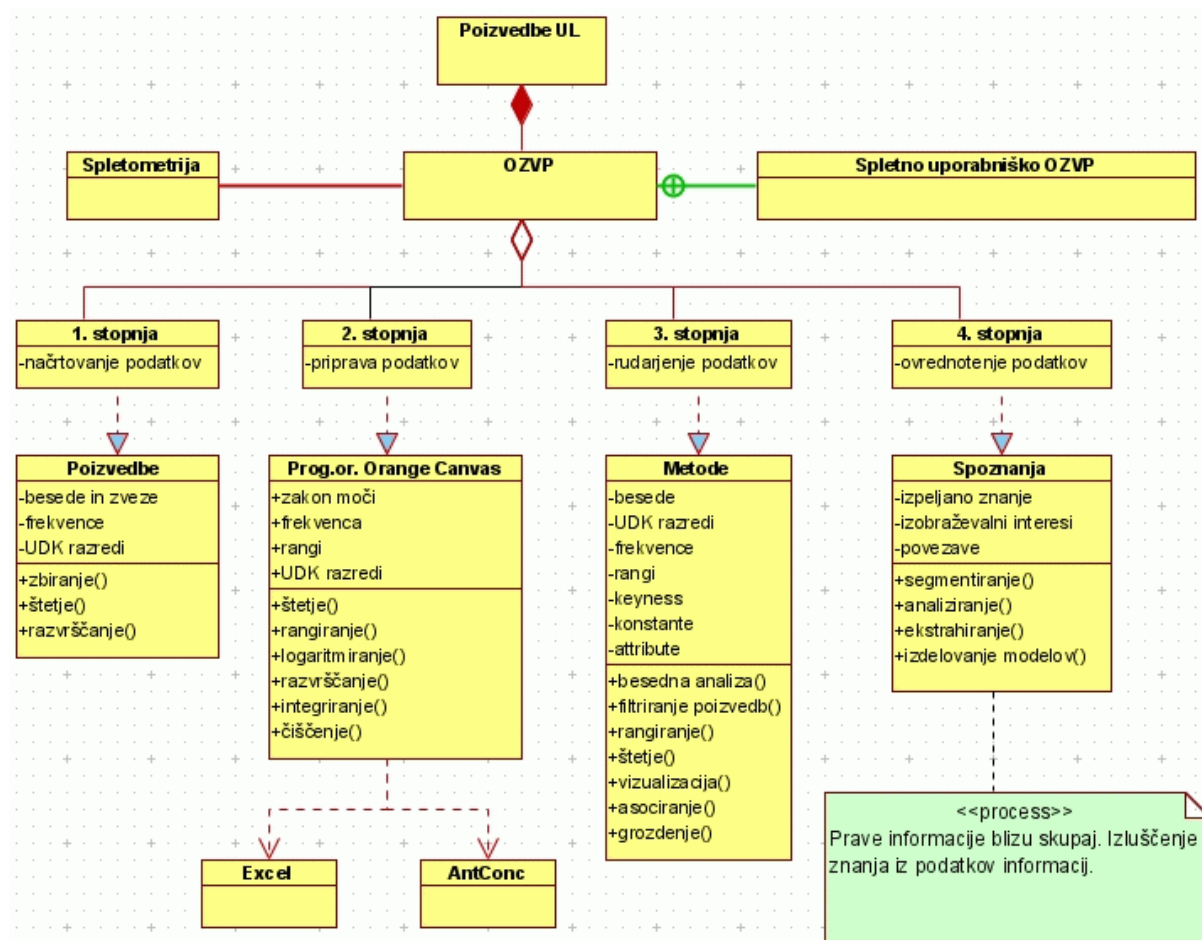
Mejnik 2 – spremljanje obiskov in poizvedb na pilotni digitalni knjižnici, nakar klasifikacija poizvedb po UDK področjih.

Mejnik 3 – ponazoritev procesov od merjenja (*MP*), zbiranja(*ZP*), obdelave/priprave (*klasifikacija*, *WME* – *webometrične analize idr.*) in analize podatkov (*TM WUM* – *odkrivanje zakonitosti v podatkih in besedilih oziroma poizvedbah*), nakar iz dobljenih novih znanj/ spoznanj profiliranje uporabnikov (*PU*).

Mejnik 4 in mejnik 5 – kakovosten model ADK, ki vključuje MRT (*ogled aplikativnega prototipa bo možen na priloženem CD-ROM v obliki prikazanega modela s pomočjo uporabe programskega orodja UML Visual Paradigm 6.0*).

Mejnik 6 – prikaže področje uporabe (*uporabniškega*) MRT v vzgojnoizobraževalnem procesu. V naslednjem podpoglavju bom predstavil postopek načrtovanja podatkov.

## 3.2 Načrtovanje podatkov – globalni pogled



### 3.2.1 Slikovni prikaz 4: Spletometrija poizvedb in postopki obdelave strukturiranih podatkov – načrtovanje podatkov

Slika 4 prikazuje na globalni ravni vse stopnje obdelave poizvedb in v nadaljevanju strukturiranih podatkov – načrtovanje podatkov. Gornji del slike 4 prikazuje hierarhični nivo ena in dva. Poizvedbe na UDK leksikonu (UL) so glavni predmet zanimanja (*z rjavo barvo zapolnjen romb – znak za kompozicijo*). S spletometričnim pristopom in metodama ter vizualizacijskimi tehnikami OZVP (*gl. ravno črto – asociativna povezava*) sem pripravil podlago za načrtovanje, pripravo, OZVP in ovrednotenje podatkov (*gl. beli romb – znak za*

*agregacijo*). Zaradi večje nazornosti naj na kratko opišem vse štiri stopnje (z *modro barvo zapolnjen trikotnik in prekinjeno črto – znak za realizacijo*) in s tem posledično temeljno vsebino drugega, tretjega, četrtega in petega poglavja tega dela:

1.) Stopnja načrtovanja – na tej stopnji sem predhodno poizvedbe sproti (*online s pomočjo notranjega spletnega iskalnika Pico Search*) zbiral in jih sproti izvozil v programsko orodje Excel. Na takšen način sem si pridobil podatke o poizvedbah in številu le teh. Prav tako sem sproti tovrstne poizvedbe, ki so jih spletni obiskovalci izvedli od 01.01.2003 do 01.01.2006, razvrščal po UDK. Pri tem razvrščanju sem si pomagal z naslednjimi klasičnimi in spletnimi priročniki:

- UDK tablice – pomožne in glavne
- Različni slovarji tujk
- Spletni UDK geselski katalog (*bo dosegljiv na CD-ROM v obliki .doc datoteke*)
- COBISS/COBIB katalog (*za razreševanje mejnih primerov*)
- Različni spletni iskalniki, še zlasti za razreševanje pojmov npr. zame nerazumljivi jeziki kot npr. španščina, italijanščina, turščina ipd. (*tovrstne poizvedbe so bile zelo redke*)
- Spletna programska orodja za spremljanje dejavnosti spletnih obiskovalcev kot npr. Nedstat, Extreme Tracking, One Stat Counter in Stat Counter.

Tako sem ustvaril naslednje kazalce: poizvedbe, frekvence in UDK razrede. Za OZVP številčnih in besednih podatkov bom uporabil tekst tabularični format (.txt), ki sem jih predhodno pretvoril iz Excelove oblike .xls, nakar jih bom potem v namen priprave podatkov ustrezno izvozil v programsko orodje Orange Canvas in AntConc. Na kratko bom poskusil tudi predstaviti programsko orodje Orange Canvas za OZVP in AntConc za OZVP v besedilih (*o tem bom poročal na strani od 48-49*). V nadaljevanju sem načrtoval naslednje faze:

2.) Stopnja priprave podatkov – pridobljene delno strukturirane podatke bom s pomočjo programskih orodij od Orange Canvas-a, Excel in AntConc pripravil tako, da si bom pridobil naslednje kazalce: ustrezno integrirane in prečiščene frekvence ter ustrezno integrirane UDK razrede, nakar bom frekvence po posameznih UDK področjih (*razen za UDK 4 – s tem bom meril poizvedovalni šum*) desetiško logaritmiral. Logaritmirane vrednosti frekvenc bom nato primerjal z rangi, ki so jih dosegale različna UDK področja. Najprej nameravam podatkovno množico z 8235 raznovrstnimi poizvedbami (*vsota vseh poizvedb je 13613*) razdeliti na tri manjše vzorce in prikazati zakon moči.

Ob ugotovitvi uspešnosti vseh treh preizkusov nameravam določiti zakon moči za vso podatkovno množico. V nadaljevanju bom dano podatkovno množico še obdelal s programskim orodjem za besedne analize AntConc. Na takšen način si bom pripravil podatke za tretjo stopnjo (*o vsem tem bom podrobneje poročal v naslednjih poglavjih*). V tej fazi si bom pridobil naslednje kazalce: konstante za posamezna UDK področja, besede, range, frekvence besed, **K** in izpeljane nove atribute. S pomočjo te stopnje si bom pridobil strukturirane podatke, ki jih bo možno obdelati s programskim orodjem Orange Canvas-om.

- 3.) Stopnja OZVP ali odkrivanja vzorcev v podatkih - na tej stopnji bom pripravljene strukturirane podatke obdelal z dvema metodama OZVP in odgovarjajočimi sedmimi vizualizacijskimi tehnikami. Pri tem bom izvedel naslednje operacije: besedna analiza poizvedb, filtriranje poizvedb, rangiranje, štetje, vizualizacija strukturiranih podatkov s pomočjo dveh metod OZVP, odkrival povezave – asociacije in grozde – clustre/skupine. Izidi oziroma izdelani vzorci podatkov mi bodo omogočili, da bom lahko le-te ovrednotil in tako izpeljal nova znanja.
- 4.) Stopnja ovrednotenja podatkov – na tej stopnji si bom na podlagi predhodnih podatkovnih vzorcev pridobil nova spoznanja in s tem nova znanja o (*izobraževalnih*) interesnih področjih spletnih obiskovalcev. Tako bom lahko analiziral, segmentiral, povezoval, ekstrahiriral kot tudi izdelal različne modele miselnih hierarhij spletnih obiskovalcev, vključno s povezanostjo med različnimi skupinami človekovega znanja (*npr. katere informacije naj bi bile blizu skupaj?*). Na podlagi tovrstnih novih odkritih znanj, bom v petem poglavju razmišljal o DK, pri čemer bom tudi izdelal model adaptivne DK ter na kratko poskušal prikazati prihodnostni razvoj le-teh.

### **3.3 Načrtovanje podatkov – osredotočen pogled**

V tem podpoglavju sem si zadal nalogo, da prvo stopnjo z nazivom "Načrtovanje podatkov" predstavim za naslednjimi vsebinskimi gradniki:

- kratka predstavitev uporabljenih programskih orodij tj. Stat Counter, Nedstat, Extreme Tracking, Pico Search (*notranji iskalnik – opcija spremljanja poizvedb in izvoz v Excel*), Orange Canvas in AntConc.
- način zbiranja podatkov
- kratek opis in določitev virov podatkov

- kratko poročilo o podatkih ( *vključuje naslednje: format, obseg ali količina, strukturiranost, število glavnih polj v tabelah in njihova poimenovanja, ključne lastnosti ipd.*)

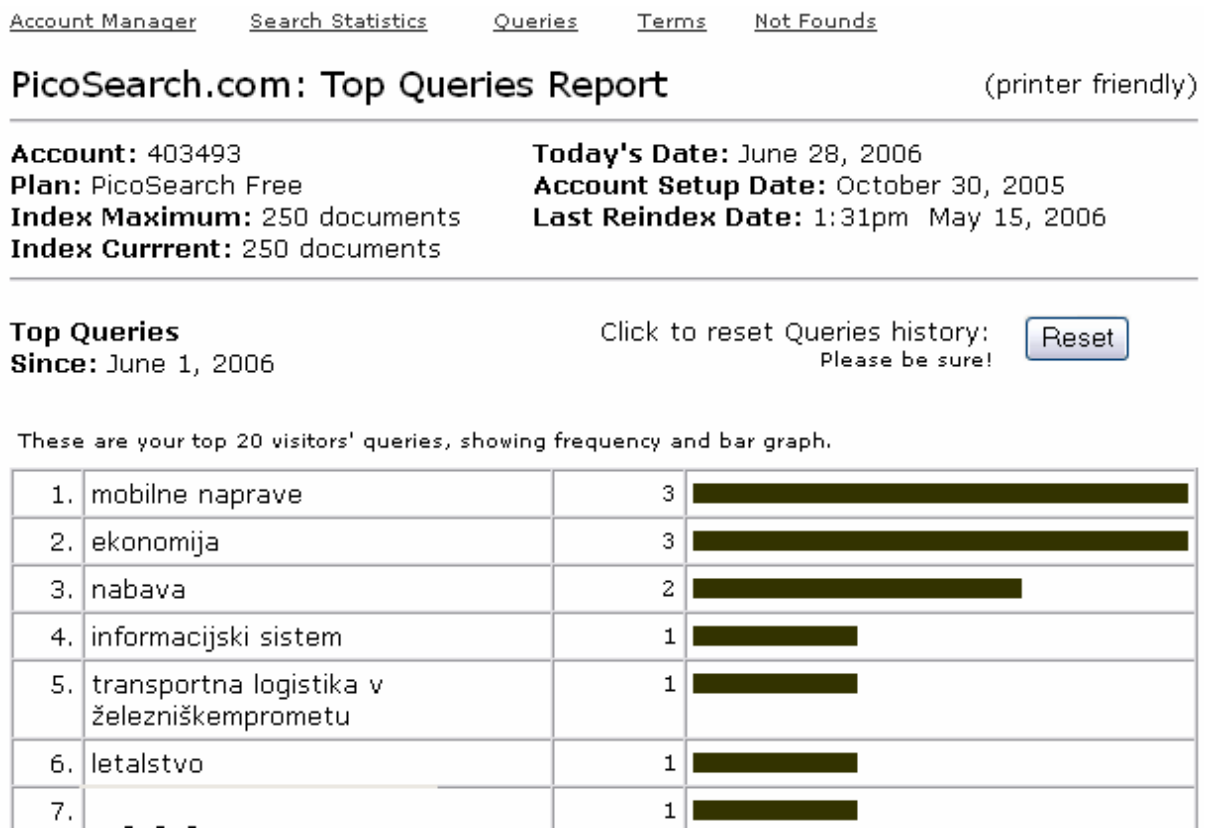
- kratko poročilo o kakovosti pridobljenih podatkih ( *vključuje: popolnost, napake, manjkajoče vrednosti, pravilnost, zanesljivost, razumljivost, relevantnost, natančnost in konsistentnost podatkov*)

### 3.3.1 Uporabljen programski orodja

Kratek opis uporabljenih programskih orodij bo osvetlil način zbiranja in v naslednji stopnji razvrščanja podatkov kot tudi prikazal izhodišče za naslednjo stopnjo tj. priprava podatkov.

### 3.3.2 Nadzorna plošča spletnega notranjega iskalnika Pico Search

V tem vpogledu bom opisal zgolj del nadzorne plošče iskalnika Pico Search tj. opcija spremljanja ali sledenja poizvedb, ki so jih izvedli spletni obiskovalci na omenjenem iskalniku.



### 3.3.3 Slikovni prikaz 5: Zajem poizvedb, ki so jih izvedli spletni obiskovalci

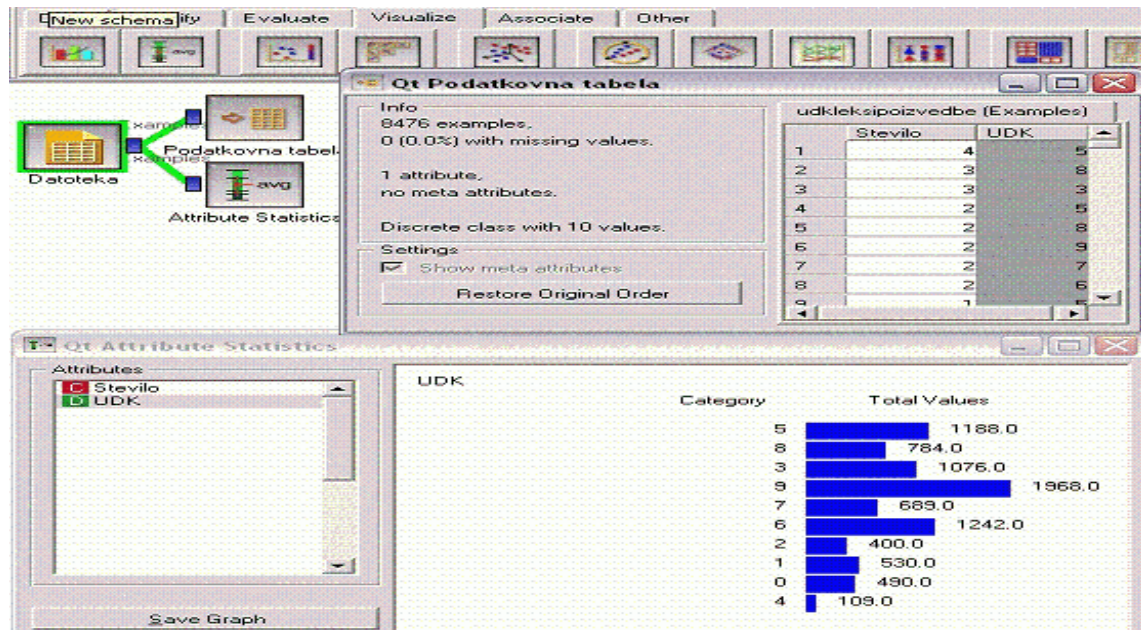
Slika 5 prikazuje del poizvedb, ki so jih izvedli spletni obiskovalci na notranjem iskalniku Pico Search. Spletno programsko orodje posreduje naslednje podatke:

- vrstni red poizvedbe
- vrsta poizvedbe (*npr. mobilne naprave*)
- število določene poizvedbe (*npr. 3 za poizvedbo mobilne naprave*)
- stolpčni diagram (*gl. črne stolpčne prikaze*)

Podatke o poizvedbah in številu sem nato izvozil v uporabniško programsko orodje Excel. V Excel-u sem potem poizvedbe sproti globalno razvrstil po UDK. V nekaterih mejnih primerih razvrščanja poizvedb sem si pomagal s spletnimi programskimi orodji za spremljanje dejavnosti spletnih obiskovalcev po straneh UDK leksikona. Z orodji kot so Stat Counter, Nedstat in Extreme Tracking sem si pridobil dodatne podatke o dejavnostih spletnih obiskovalcev (*ob tem naj ponovno omenim, da sem ta dodaten pristop uporabil zgolj v primerih manj razumljivih in mejnih poizvedb. Ta način spremljanja ni bilo potrebno pogosto uporabiti*). S tovrstnimi pristopi sem si pridobil naslednje podatke o spletnih obiskovalcih:

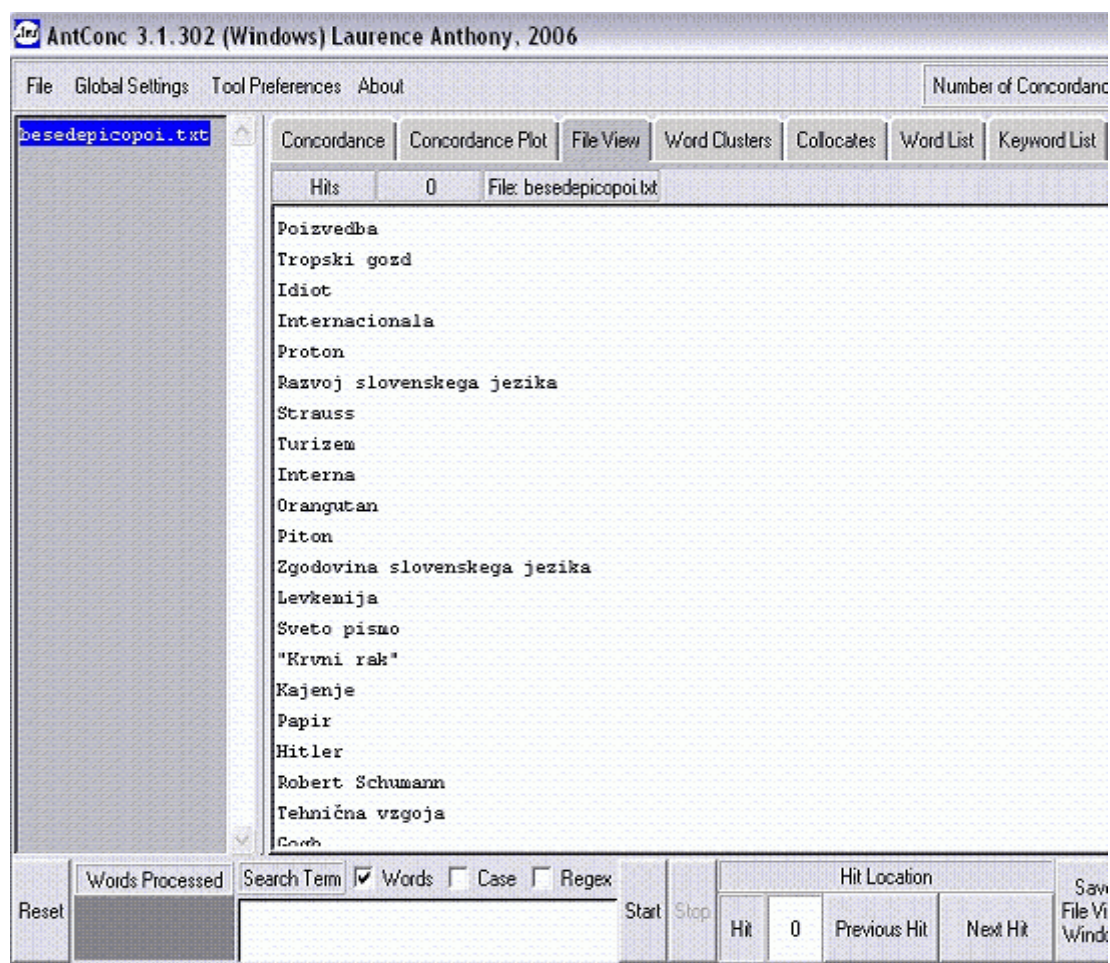
- IP (*Internet Protocol*) naslov in oskrbovalca internetnih storitev za določenega spletnega obiskovalca (*npr. siol.si*).
- Od kod je prispel spletni obiskovalec (*npr. z iskalnika Google*)
- Seznam najbolj približanih spletnih strani
- Uporabljene ključne besede s strani spletnih obiskovalcev, ki so jih preko iskalnikov, katalogov, virtualnih vodičev ipd. napotili na glavno stran UDK leksikona
- Dolžina spletnih sej spletnih obiskovalcev
- Podatke o poteh brskanja s strani spletnih obiskovalcev

### 3.3.4 Orange Canvas in Antconc



### 3.3.5 Slikovni prikaz 6: Programsko orodje za OZVP Orange Canvas

Slika 6 prikazuje del uporabniškega vmesnika programskega orodja Orange Canvas za OZVP. O tem programskem orodju je možno najti podroben opis na spletnih straneh AILAB.si, zato ocenjujem, da bom v tem delu to programsko orodje predstavil zgolj v znamenju namena uporabe, kajti z Orange Canvas-om nameravam v tej prvi stopnji načrtovanja podatkov izdelati izhodišče za kasnejšo pripravo podatkov. Excel-ovo datoteko (.xls) bom najprej pretvoril v tekst tabularično datoteko (.txt). V nadaljevanju bom (gl. slika 6) naložil .txt datoteko v znak za datoteko in s tem posledično v podatkovno tabelo (gl. slika 6). Znak "Datoteka" bom nato povezal z znakom "Statistika atributov (Attribute Statistics)." V fazi načrtovanja podatkov imam še opraviti z manj strukturiranimi podatki, ki jih želim predelati v strukturirane podatke, kar bom šele lahko uresničil v drugi stopnji tj. priprava podatkov. S tem programskim orodjem načrtujem predelavo manj strukturiranih številčnih podatkov v strukturirane (gl. slika 6 podatki o številu in po UDK razvrščenih poizvedb).



### 3.3.6 Slikovni prikaz 7: Programsko orodje AntConc za OZVP v besedilih

Slika 7 prikazuje uporabniško okensko okolje programskega orodja AntConc 3.1.302 za OZVP v besedilih in del naložene .txt datoteke s poizvedbami. S programskim orodjem AntConc bom v prvi fazi načrtoval predelavo manj strukturiranih tekstovnih podatkov v strukturirane. V poznejši stopnji priprave podatkov, bo potrebno podatke še prečistiti ter integrirati z že omenjenim statističnim kazalcem **K**. Ocenjujem, da podrobnejša predstavitev tovrstnega programskega orodja tudi v tem primeru ni potrebna, kajti izkušeni bralci tega dela (v primeru, da si to želijo) si lahko podrobnosti o AntConc ogledajo na domači strani prof. dr. Laurence Anthony-a. Povrh teh dveh programskih orodij nameravam kot dodaten pripomoček uporabiti še programski orodji kot sta Ilog Discovery (tj. za vizualizacijo številčnih in besednih podatkov) in CBA 1.0 (tj. za odkrivanje asociacijskih pravil).

### 3.4 Določitev virov podatkov in opis

Podatke, ki sem jih zajel s pomočjo spletnega programskega orodja Pico Search so izid poizvedb spletnih obiskovalcev na notranjem iskalniku, ki je vgrajen na glavni strani UDK leksikona. Gre za pridobljene zunanje ali eksterne podatke, ki so na tujem strežniku shranjeni v .HTML obliki.

Z ozirom na dejstvo, da sem ves čas od 2003 do 2006 uporabil brezplačno različico, sem moral večkrat na dan spremljati izide poizvedb, kajti brezplačna različica zabeleži zgolj 20 različnih poizvedb, nakar se vse nadaljnje prepisejo. Kljub tej oviri, nisem imel težav pri zajemu tovrstnih podatkov, kar pomeni, da nisem v tem obdobju izgubil niti enega podatka (*v povprečju je bilo dnevno okoli 15 poizvedb*). Prav tako je omenjeni tuji strežnik v tem obdobju deloval brez hujših motenj, tako da je bila storitev poizvedovanja spletnim obiskovalcem vedno na voljo. Zajete sledi oziroma poizvedbe spletnih obiskovalcev so bile s časovnega vidika vedno aktualne in z verodostojnega vira. Podatkovni viri oziroma zajete poizvedbe so bile v obliki besednih zvez (*enobesedne, dvobesedne in večbesedne*), števil (UDK vrstilci idr. števil), matematičnih znakov (npr. +, - ipd.) in interpunkcijskih znamenj (npr. , . :/"). V večini primerov sem imel pri poizvedbah opraviti z osrednjim evropskim znakovnim naborom in zelo redko z drugimi (*npr. zahodno evropski, cirilica*). Pretežni del poizvedb je bilo v slovenskem jeziku in zelo redke so bile poizvedbe v tujih jezikih (*kot npr. v nemščini, angleščini, španščini, portugalščini, norveščini, turščini*).

### 3.5 Kratko poročilo o podatkih

- format: izvoz .html datoteke v .xls obliko (*zbiranje podatkov v Excel-u*) in v namen kasnejše stopnje priprave podatkov sem .xls datoteko pretvoril .txt datoteko.
- obseg: zbranih 13613 poizvedb, od tega 8235 raznovrstnih
- strukturiranost: zajeti podatki so bili delno strukturirani, kajti vrednosti nekaterih poizvedb, ki so se večkrat ponavljale niso bile seštete in so se obravnavale zaporedno. Prav tako je bilo razvrščanje po UDK zaporedne narave.
- Število glavnih poimenovanj polj in ključne lastnosti v Excelovi tabeli: vrstni red, besedne zveze (*poizvedbe*), število in ustrezno dodeljeni UDK vrstilci za poizvedbe. Atributi so poizvedbe, število in UDK razredi.<sup>29</sup>

---

<sup>29</sup> Pri načrtovanju podatkov sem se nekoliko vzgledoval po naslednjem spletnem delu: <http://www.crisp-dm.org/CRISPWP-0800.pdf> (2006-06-29)

### 3.6 Kratko poročilo o kakovosti pridobljenih podatkov

Popolnost: pridobljeni podatki so bili popolni.

Napake: morebitne napake ali nedoslednosti pri UDK razvrščanju sem sproti popravljaj.

Manjkajoče vrednosti: manjkajočih vrednosti ni bilo

Pravilnost: pri interpretaciji nekaterih poizvedb sem sproti izvajal rekonstrukcijo le-teh, tako da bi s tem postopkom doumel pravo namero spletnih obiskovalcev.

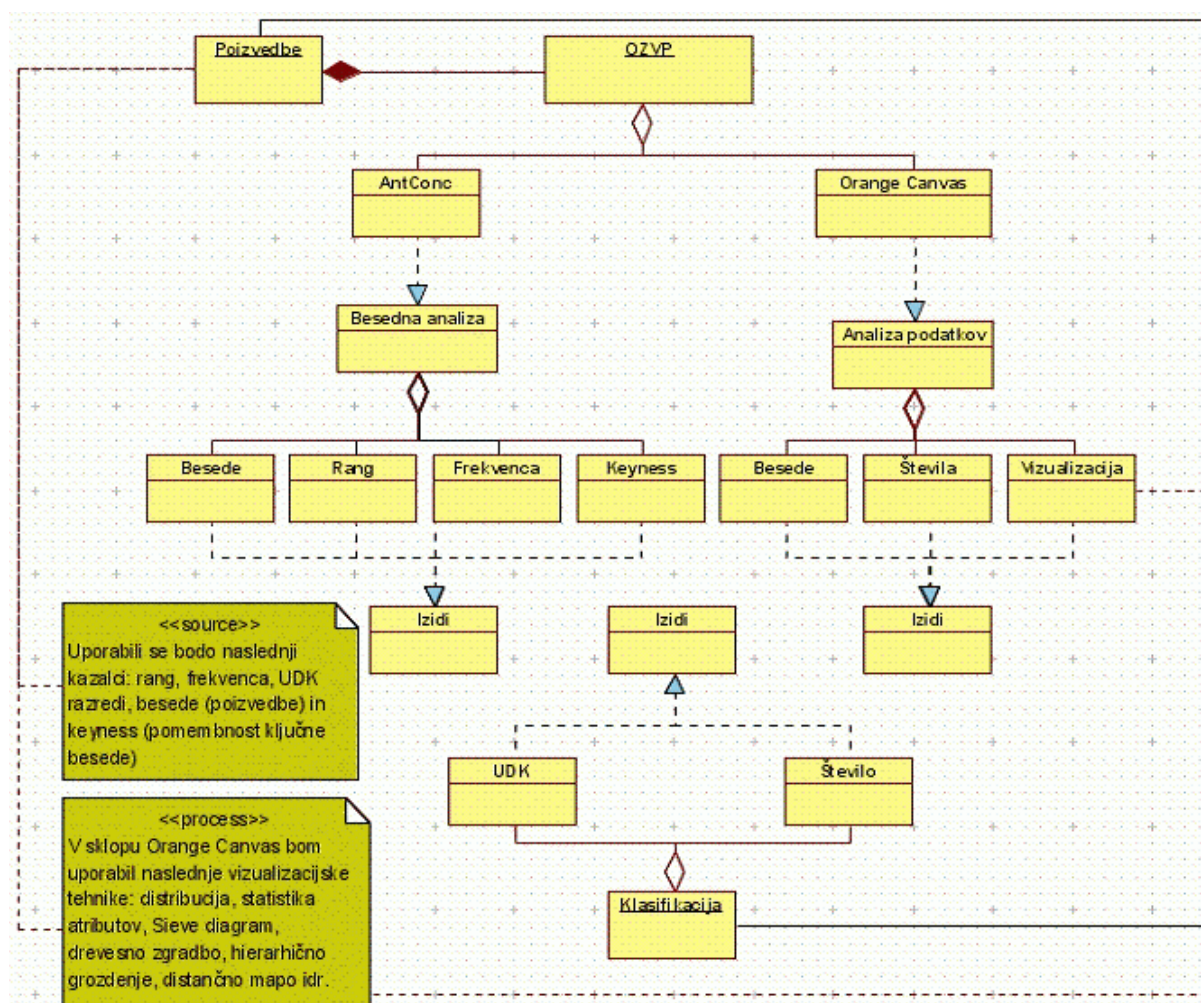
Zanesljivost: spletno programsko orodje Pico Search je delovalo povsem zanesljivo, tako da so tudi zajeti podatki zanesljivi.

Razumljivost: redke so bile tiste poizvedbe, ki so bile manj ali celo nerazumljive. V takšnih primerih sem tovrstne poizvedbe razvrstil v področje UDK 4 (*od 13613 poizvedb je bilo 152 poizvedb manj razumljive narave in sem jih označil kot poizvedvalni šum*). S tega izhajajoč so bili podatki v skoraj 99% primerih relevantni in uporabni.

Konsistentnost: zajeti podatki si med seboj niso bili v nasprotju.

Na podlagi stopnje načrtovanja podatkov, sem si zdaj pripravil vse potrebno za naslednjo stopnjo tj. priprava podatkov, ki je predpogoj za kasnejše odkrivanje zanimivih vzorcev v podatkih. Odkrivanje zanimivih vzorcev v podatkih dosežemo pa z uporabo ustreznih metod, algoritmov in vizualizacijskih tehnik. Prav o tem bom podrobneje poročal v nadaljevanju!

## 3.7 Priprava podatkov



### 3.7.1 Slikovni prikaz 8: Globalni pogled na pripravo podatkov za OZVP

Slika 8 z globalnega vidika prikazuje pripravo podatkov za odkrivanje zanimivih vzorcev v podatkih oziroma za še kasnejšo OZVP. Ta globalni pogled na pripravo podatkov za OZVP je razdeljen na tri stopnje:

1. stopnja ali klasifikacija poizvedb: poizvedbe (*gl. neprikinjeno črto od poizvedbe do klasifikacije – asociacija*) se razvrščajo po UDK razredih in po številu. S pomočjo programskega orodja Orange Canvas te poizvedbe pripravimo tako, da dobimo kot izid strukturirane številčne podatke, kajti s pomočjo nalaganja ustrezno pripravljene .txt datoteke in ustvarjene povezave na statistične attribute (*Attribute Statistics*) dobimo na vpogled vsoto poizvedb po posameznih UDK razredih. Tovrstno vsoto poizvedb znotraj posameznih UDK razredov sem potem s pomočjo programskega orodja Excel (*Data – Autofilter in ukaz SUM*) določil tako, da sem dobil dejansko frekvenčno zastopanost po

posameznih UDK razredih. Število vseh poizvedb je 13613, število raznovrstnih poizvedb je 8235, a različnih vrednosti frekvenc (N) je 17.

$N = \{1 (4865), 2 (2320), 3 (633), 4 (248), 5 (84), 6 (50), 7 (20), 8 (16), 9 (6), 10 (8), 11 (2), 12 (2), 13 (2), 14 (1), 15 (1), 16 (1), 22 (1)\}$ . Znotraj okroglih oklepajev so vrednosti pogostosti posameznih razredov frekvenc. V naslednji preglednici bom po rangih prikazal vrednosti frekvenc po posameznih UDK razredih in dejansko oziroma skupno vrednost frekvenc po posameznih UDK razredih.

### 3.7.2 Preglednica 2: Rang, frekvence in logaritmirane frekvence znotraj posameznih UDK razredov

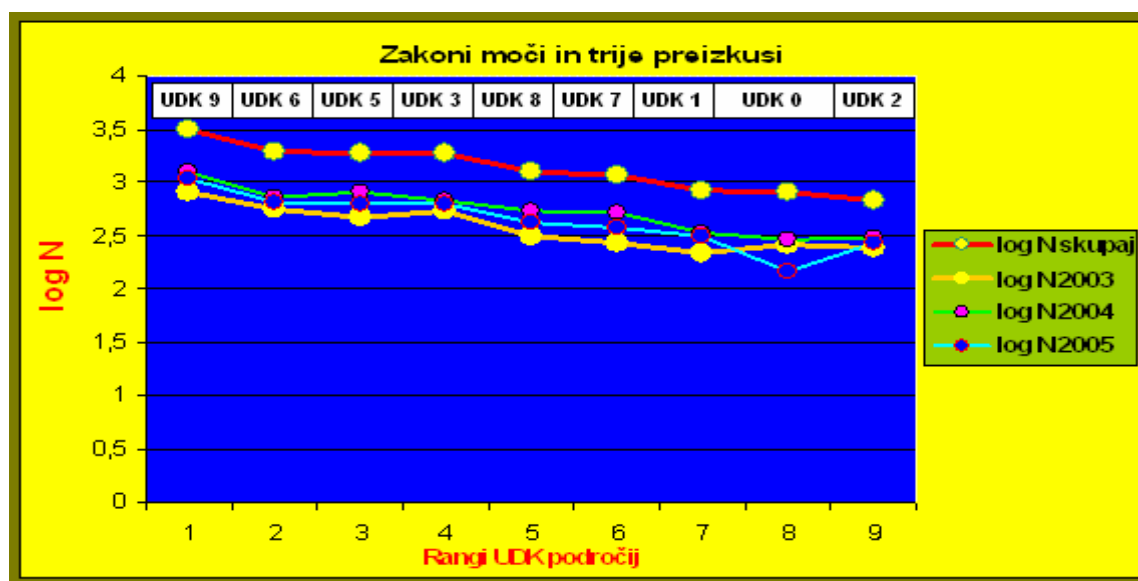
Rang (R)	UDK razredi	N skupaj z UDK razredi	log N	N UDK razredov
1	9	3127	3,4951	1890
2	6	1917	3,2826	1200
3	5	1863	3,2702	1155
4	3	1828	3,2620	1051
5	8	1256	3,0990	762
6	7	1155	3,0626	678
7	1	846	2,9274	519
8	0	809	2,9079	476
9	2	660	2,8195	395
10	4	152	X	109

Preglednica 2 prikazuje range UDK razredov, njihove skupne ter posamične frekvence (N) in logaritmirane N-je ( $\log N$ ). Prav frekvence, njihove desetiške logaritme in range bom v poznejši stopnji uporabil za določevanje zakona moči in izračunavanje konstant za posamezna UDK področja. Razvrščene poizvedbe s področja UDK 4 bom uporabil za izračunavanje poizvedovalnega šuma, zato odgovarjajoča frekvenca tudi ni desetiško logaritmirana (gl. gornjo preglednico znak X). Preden bom lahko prikazal porazdelitev zakona moči, bo pred tem potrebno še pripraviti tri preizkuse. To pomeni, da je potrebno dani podatkovni vzorec razdeliti na tri enake dele oziroma vzorce, nakar podatkom štejemo frekvence znotraj posameznih UDK razredov, frekvence desetiško logaritmiramo in rangiramo. Izide potem prikažemo slikovno (na x osi so rangi, na y osi so  $\log N$ ). S

temi preizkusi želimo ugotoviti ali lahko podatkovni množici določimo zakon moči, kar pomeni, da kažejo izidi po različnih časovnih obdobjih sorazmerno neodvisnost od časa. V primeru, da bodo velika odstopanja, potem prej omenjeno metodo ne bo možno uporabiti, ampak se bo potrebno posluževati analizi trenda, kar pomeni časovno odvisnost.

### 3.7.3 Preglednica 3: Rangirana UDK področja in logaritmirane frekvence za leto 2003, 2004, 2005 in za vse skupaj

Rang (R)	log N skupaj	UDK	Rang (R)	log N 2003	Rang (R)	log N 2004	Rang (R)	log N 2005
1	3,4951	9	1	2,8971	1	3,1000	1	3,0354
2	3,2826	6	2	2,7419	3	2,8645	2	2,8122
3	3,2702	5	4	2,6590	2	2,9053	4	2,7917
4	3,2620	3	3	2,7300	4	2,8299	3	2,8000
5	3,0990	8	5	2,4942	5	2,7348	5	2,6201
6	3,0626	7	6	2,4330	6	2,7168	6	2,5717
7	2,9274	1	9	2,3385	7	2,5211	7	2,4941
8	2,9079	0	7	2,4116	9	2,4624	9	2,1553
9	2,8195	2	8	2,3856	8	2,4713	8	2,4330
X	X	4	X	X	X	X	X	X



### 3.7.4 Slikovni prikaz 9: Zakon moči in preizkusi

Preglednični prikaz 3 in slika 9 ponazarjata zakon moči po posameznih obdobjih za leta 2003, 2004, 2005 in za vsa leta skupaj. Tako preizkusi kot tudi celotni vzorec dajejo kot izid sorazmerno ravno premico, kar kaže na neko stabilno obnašanje preučevanega sistema tj. poizvedovanje spletnih obiskovalcev na notranjem iskalniku UDK leksikona. Z metodo zakona moči sem dokazal, da gre pri preučevanju poizvedb spletnih obiskovalcev na omenjeni izvedbi za odprt sistem, katerih dinamik so kvalitativno gledano (*na veliki skali*) sorazmerno neodvisni od časa (*npr. razni sezonski pojavi*). Ko primerjamo med sabo range in logaritmirane frekvence poizvedb po različnih UDK področjih bomo vedno dobili kot izid neko značilno obnašanje sistema. Na ta sistem vplivajo različni dejavniki kot so spletni obiskovalci, urednik spletnega sistema, podatki/informacije, oskrbovanost z informacijami, spletne izvedbe kot npr. notranji iskalnik, informacijske potrebe ljudi, komunikacijska oprema, jezik na spletnem sistemu in nenazadnje, kar je tudi zelo pomembno t.i. zunanji svetovni iskalniki, med njimi še zlasti izstopata dva, to sta Google in Najdi.si. Oba iskalnika predstavljata glavna vhoda v spletni sistem UDK leksikona. Odveč bi bilo sedaj pri pripravi podatkov za naslednje poglavje nadrobno opisati oba prikaza, ker sta dovolj izpovedna, vendar ocenjujem, da ne bo odveč, če bom nekoliko podrobneje opisal izide rangov UDK področij in desetiško logaritmiranih frekvenc za celotno obdobje od 2003 do 2006.

Poizvedbe, ki sem jih razvrstil pod UDK 9 je bilo največ, zato je to področje doseglo prvo mesto ali rang 1 s 3127 poizvedbami ( $\log N = 3,4951$ ). Rang 2 je doseglo področje UDK6 s 1917 poizvedbami ( $\log N = 3,2826$ ), medtem ko je področje UDK 5 doseglo rang 3 s 1863 poizvedbami ( $\log N = 3,2702$ ). V vrhno skupino spada tudi področje UDK 3, ki je doseglo rang 4 s 1828 poizvedbami ( $\log N = 3,2620$ ). V srednjo skupino glede pogostosti poizvedb v celotnem obdobju lahko izpostavim področje UDK 8 (*rang 5, 1256 poizvedb,  $\log N = 3,0990$* ) in področje UDK 7 (*rang 6, 1155 poizvedb,  $\log N = 3,0626$* ). V spodnjo skupino z najmanjšim številom poizvedb lahko uvrstim področje UDK 1 (*rang 7, 846 poizvedb,  $\log N = 2,9274$* ), področje UDK 0 (*rang 8, 809 poizvedb,  $\log N = 2,9079$* ) in na zadnje mesto se je uvrstilo področje UDK 2 tj. rang 9 in s 660 poizvedbami ( $\log N = 2,9079$ ). Področje UDK 4 je kot že omenjeno, po mednarodnem sistemu UDK prazno in zaenkrat temu še niso dodelili pomena oziroma uporabe. Poizvedbe, ki sem jih razvrstil v področje UDK 4 so takšne narave, da jih ni možno v nobeno področje človekovega znanja razvrstiti in zato predstavljajo nekakšen poizvedovalni (*izhodni*) šum. Prav zaradi tega tovrstne poizvedbe nisem vključil v meritev zakona moči. Poizvedb s področja UDK 4 je

bilo vsega skupaj 152 ali slabih 1,12 % vseh poizvedb tj. 13613. Pri poznejšem OZVP pa bom tudi ta podatek uporabil.

Na podlagi izvedenih meritev bom kot zanimivost še pripravil izračunane konstante za posamezna UDK področja in pri tem še vključil področje UDK 4. V ta namen bom uporabil že znani matematični obrazec s strani 12 tj.  $C = N \cdot r$  ( $C$  ... konstanta,  $N$  ... frekvenca,  $r$  ... rang)

$$C_{UDK\ 9} = N_{UDK\ 9} \cdot r_{UDK\ 9} = 3127 \cdot 1 = 3127 \dots 22,97 \%$$

$$C_{UDK\ 6} = N_{UDK\ 6} \cdot r_{UDK\ 6} = 1917 \cdot 2 = 3834 \dots 14,08 \%$$

$$C_{UDK\ 5} = N_{UDK\ 5} \cdot r_{UDK\ 5} = 1863 \cdot 3 = 5589 \dots 13,68 \%$$

$$C_{UDK\ 3} = N_{UDK\ 3} \cdot r_{UDK\ 3} = 1828 \cdot 4 = 7312 \dots 13,43 \%$$

$$C_{UDK\ 8} = N_{UDK\ 8} \cdot r_{UDK\ 8} = 1256 \cdot 5 = 6280 \dots 9,23 \%$$

$$C_{UDK\ 7} = N_{UDK\ 7} \cdot r_{UDK\ 7} = 1155 \cdot 6 = 6930 \dots 8,48 \%$$

$$C_{UDK\ 1} = N_{UDK\ 1} \cdot r_{UDK\ 1} = 846 \cdot 7 = 5922 \dots 6,21 \%$$

$$C_{UDK\ 0} = N_{UDK\ 0} \cdot r_{UDK\ 0} = 809 \cdot 8 = 6472 \dots 5,94 \%$$

$$C_{UDK\ 2} = N_{UDK\ 2} \cdot r_{UDK\ 2} = 660 \cdot 9 = 5940 \dots 4,85 \%$$

$$C_{UDK\ 4} = N_{UDK\ 4} \cdot r_{UDK\ 4} = 152 \cdot 10 = 1520 \dots 1,12 \%$$

$C_p = \Sigma C_{UDK} / 10 = 5292,6$  je izid konstante za vso človekovo znanje oziroma za vsa UDK področja, vključno s področjem UDK 4, ki sem ga opredelil kot poizvedovalni (izhodni) šum. Odstotkovne deleže za posamezna UDK področja sem zabeležil za izidi posameznih konstant. V tem podglavju bom še navedel razlike med frekvencami po UDK področjih in pogostosti pojavljanja UDK področij brez frekvenc po naslednjem matematičnem obrazcu:  $R_{UDK} = N_{UDK} - UDK$ .

$$R_{UDK\ 9} = N_{UDK\ 9} - UDK\ 9 = 3127 - 1890 = 1237; R_{UDK\ 6} = N_{UDK\ 6} - UDK\ 6 = 1917 - 1200 = 717; R_{UDK\ 5} = N_{UDK\ 5} - UDK\ 5 = 1863 - 1155 = 708; R_{UDK\ 3} = N_{UDK\ 3} - UDK\ 3 = 1828 - 1051 = 777; R_{UDK\ 8} = N_{UDK\ 8} - UDK\ 8 = 1256 - 762 = 494; R_{UDK\ 7} = N_{UDK\ 7} - UDK\ 7 = 1155 - 678 = 477; R_{UDK\ 1} = N_{UDK\ 1} - UDK\ 1 = 846 - 519 = 327; R_{UDK\ 0} = N_{UDK\ 0} - UDK\ 0 = 809 - 476 = 333; R_{UDK\ 2} = N_{UDK\ 2} - UDK\ 2 = 660 - 395 = 265, R_{UDK\ 4} = N_{UDK\ 4} - UDK\ 4 = 152 - 109 = 43.$$

Na podlagi izidov teh enostavnih izračunov je možno opaziti določeno zanimivost, kajti rangi so se pri nekaterih UDK področjih spremenili (npr. UDK 3, UDK 1, UDK 0).

### 3.8 Druga stopnja ali priprava besedilnega dela podatkov

Na tej stopnji bom besedilni del podatkov (*gl. slikovni prikaz 8*) oziroma poizvedbe pripravil tako, da bom kot izid dobil kazalce kot so frekvence, range in pomembnost ali moč ključnih besed. Izvirne poizvedbe bodo pretvorjene v besede, ki bodo prav tako razvrščena v UDK področja. Pri tovrstni pripravi podatkov si bom pomagal z že omenjenim programskim orodjem AntConc 3.1.302, ki se uporablja za besedne analize oziroma za OZVP v besedilih. Izvirno Excel-ovo zbirko poizvedb bom pretvoril v .txt datoteko in jo shranil, nakar bom le-to uvozil v programsko orodje AntConc. V naslednjem koraku bom določil (*predhodno še moram določiti centralno evropsko jezikovno nastavitvev CP 1250*) oziroma sprožil opcijo pogled datoteke (*File view*) in kliknil v levem oknu direktorija ime datoteke, v tem primeru je to antconc0306.txt. V glavnem oknu se nam bodo prikazale vse poizvedbe po izvirnem vrstnem redu. Zdaj moram določiti glavno opcijo »Besedni seznam (*Word list*)«, znotraj tega »Razvrsti po številu (*Sort by frequency*)«, nakar pritisnemo opcijo »Začni (*Start*)«, nakar dobimo na vpogled besede, ki so razrščene po frekvencah in rangih. Z ozirom na to, da se zdaj v tej besedni zbirki nahaja veliko takšnih besed, ki se sicer lahko zelo pogosto pojavljajo (*npr. za, in, v ipd.*), vendar so pa s pomenskega pogleda manj izrazita kot je temu slučaj pri drugih besedah (*npr. zgodovina, stres, življenjepis, Bach*), je potrebno besede z manjšo stopnjo izrazitosti očistiti oziroma izločiti. To izvedemo tako, da najprej sprožimo gornji navigacijskih gumb »Prednostno orodje (*Tool preferences*)« in znotraj tega določimo opcijo »Prednostni seznam ključnih besed (*Keyword list preferences*)«, v katerem določimo podopcije kot so rang, frekvenco, ključno besedo, **K** (*Keyness*), statistično operacijo Log – likelihood in povprečna vrednost (*Average Value*). V nadaljevanju izberemo antconc0306.txt in sprožimo gumb »vredno (OK)«. Zdaj je vse pripravljeno za naslednji korak. Odpremo opcijo »Seznam ključnih besed (*Keyword list*)« in odpre se nam novo okensko okolje. V spodnjem okencu direktorija določimo »Razvrsti po **K** (*Sort by Keyness*) in obkljukamo majhen prazen kvadratik »Primer (*Case*)«, nakar ponovno sprožimo gumb »Začni (*Start*)«.

Rank	Freq	Keyness	Keyword
1	91	68.648	zgodovina
2	53	54.236	stres
3	31	42.975	bach
4	30	41.589	udk
5	28	38.816	kaj
6	26	36.044	vivaldi
7	25	34.657	haydn
8	24	33.271	beethoven
9	24	33.271	napoleon
10	23	31.885	shakespeare
11	20	27.726	anton
12	20	27.726	seliškar
13	19	26.340	tone
14	16	22.181	bonaparte

### 3.8.1 Slikovni prikaz 10: Čiščenje poizvedb s pomočjo programskega orodja AntConc

Slika 10 ponazarja očiščene poizvedbe s pomočjo programskega orodja AntConc. Kot izid dobimo na vpogled ključne besede, ki so razvrščene po rangih, frekvencah in **K**. Že s tega vpogleda je razvidno, da so pomensko manj izrazite besede že izločene. Vendar je potrebno določene besede, za katere vemo, da so pomensko manj izrazite kot npr. vezniki, predloge idr. še očistiti in s tem posledično je potrebno besedam z večjo stopnjo pomena na novo določiti range in UDK vrstilce. Prav to izvedemo tako, da podatke izvozimo v Excel, nakar podatke ustrezno obdelamo in jih ponovno shranimo v .txt datoteko. Vzporedno temu v AntConc z opcijo »Besedni grozd (*Word Cluster*)« še preverimo vsebinski kontekst besed, tako da jim lahko pravilneje določimo UDK vrstilce. Tako pripravljene strukturirane podatke je potem možno še naprej obdelati oziroma vizualizirati s programskim orodjem Orange Canvas, s čimer se bom ukvarjal v naslednjem poglavju.

### 3.8.2 Opis podatkov<sup>30</sup>

Po besedni analizi in postopkom prvega in drugega čiščenja podatkov sem kot izid dobil naslednje podatke:

- rang besede (*skupaj 1155 vrednosti, od tega 80 raznovrstnih*)

<sup>30</sup> Zbirke podatkov bodo dosegljive v spremnem gradivu doktorskega dela na CD-ROM.

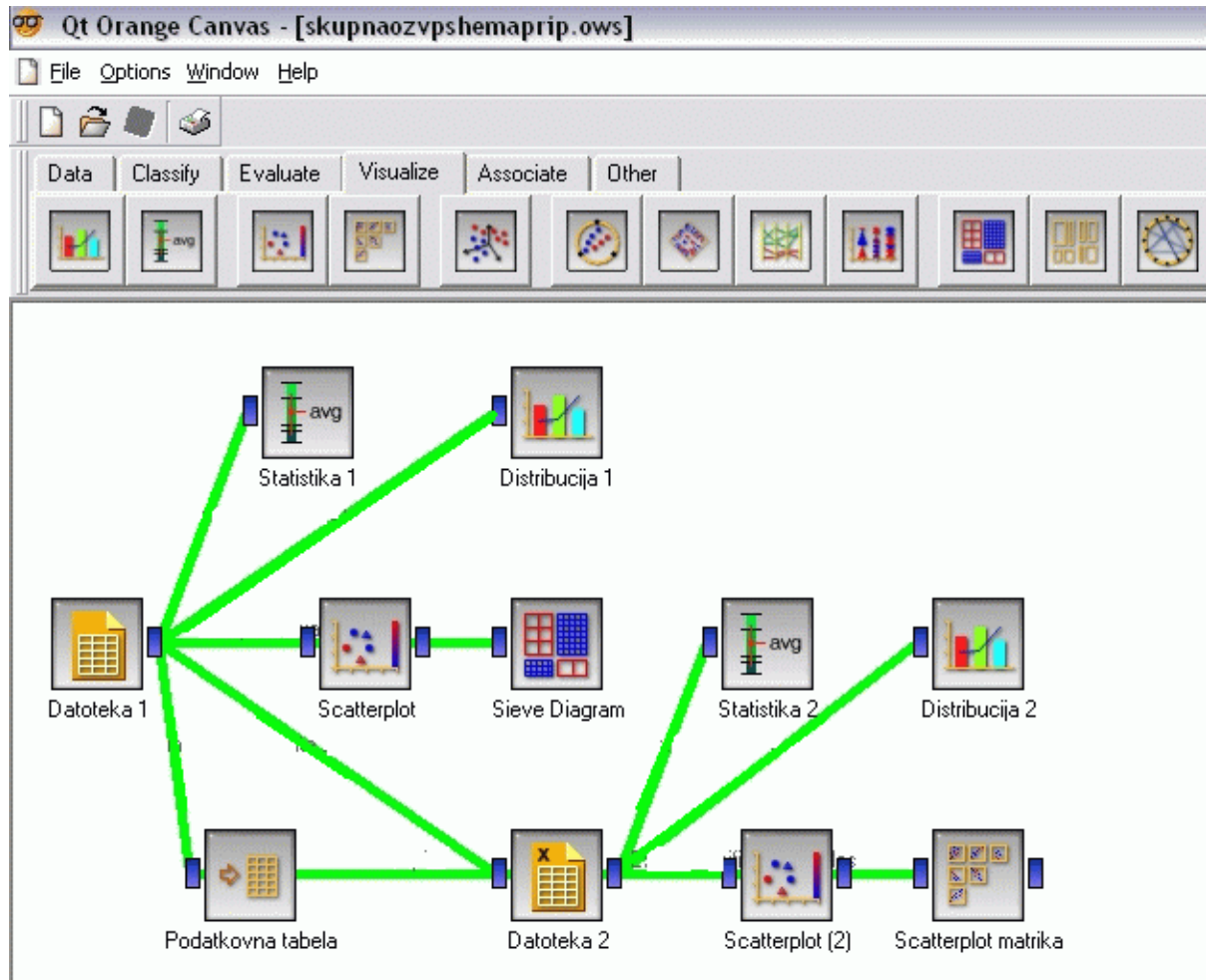
- frekvenca besede (*skupaj 1155 vrednosti, od tega 32 raznovrstnih*)
- **K** (*skupaj 1155 vrednosti, od tega 80 raznovrstnih*)
- Besede (*1155 raznovrstnih besed*)
- UDK področja (*pogostost pojavljanja brez upoštevanja frekvenc*): UDK 9 = 449, UDK 5 = 136, UDK 6 = 120, UDK 3 = 115, UDK 8 = 101, UDK 7 = 78, UDK 0 = 59, UDK 2 = 50, UDK 1 = 47 in UDK 4 = 0. S čiščenjem podatkov sem tako odstranil tudi tiste poizvedbe, ki sem jih prej razvrstil kot UDK 4, kajti tovrstne poizvedbe so bile zelo redke in povrh tega niso nikoli bile vsebinsko povezane z določeno drugo poizvedbo, kar sem ugotovil z besedno analizo poizvedb. Ko upoštevamo še zraven vrednosti frekvenc za posamezna UDK področja dobimo naslednji izid: UDK 9 = 2186, UDK 5 = 491, UDK 7 = 389, UDK 8 = 374, UDK 3 = 351, UDK 6 = 293, UDK 0 = 284, UDK 1 = 205, UDK 2 = 191 in UDK 4 = 0. Opazimo lahko, da se izidi nekoliko razlikujejo, kajti še zlasti frekvenca področja UDK 6 je nižja od frekvenc UDK 9, UDK 5, UDK 7, UDK 8 in UDK 3 in se tako uvršča šele na šesti rang. Ta izid je posledica analize moči ali pomembnosti ključnih besed, kajti predstavniki besed s področja UDK 6 so dosegli nižje vrednosti **K**.

### 3.8.3 Kratko poročilo o čiščenju podatkov

Po prvem postopku čiščenja podatkov s programskim orodjem AntConc sem izločil vse tiste besede, ki so se pri poizvedbah pojavile manj kot dvakrat. Povrh tega so bile izločene tudi tiste besede, ki jih ni bilo možno vsebinsko povezati z drugimi (*nizka konkordanca*) in da niso bile povezane v grozd z drugimi besedami (*besedni grozdi*). Izločene so bile tudi tiste besede, ki so imele nizke vrednosti **K**. Kljub temu so v zbirki besed ostale štiri besede (*kaj, kako, h, zakaj*), ki so dosegle sorazmerno visoke **K**-je. Pri drugem čiščenju sem tudi te podatke očistil, ker menim, da njihov pomen ni ključen za poznejše OZVP. Prav tako bi tudi izločil npr. veznike ipd., če bi se v tej zbirki besed pojavili. V novi zbirki besed so ostale tiste, ki bodo za poznejše OZVP ključnega pomena. Vse besede sem ponovno globalno razvrščal po UDK, tako da sem kot izid dobil naslednje množice, ki jih bom v nadaljevanju proučeval: rang, frekvenca, **K**, besede in UDK. S pomočjo teh pripravljenih podatkov bom lahko v naslednjem poglavju odkrival asociacije med UDK področij kot tudi hierarhične grozde, kar mi bo dodatno olajšalo segmentacijo spletnih obiskovalcev v zvezi z (*izobraževalnimi*) področnimi zanimanji. Tovrstne strukturirane podatke bom lahko v obliki .txt datoteke uvozil v programsko orodje Orange Canvas in le-te podatke povezal s podatki, ki so izid prve stopnje priprave podatkov. V

naslednjem podpoglavju bom še slikovno in opisno predstavil tretjo stopnjo priprave podatkov s programskim orodjem Orange Canvas.

### 3.9 Tretja stopnja priprave podatkov



#### 3.9.1 Slikovni prikaz 11: Priprava podatkov s pomočjo programskega orodja Orange Canvas

Slika 11 prikazuje prvi del tretje stopnje priprave podatkov, ki pomeni izhodišče za kasnejšo OZVP. V tej osnovni shemi so zajeti strukturirani podatki, ki so izid prve in druge stopnje priprave podatkov in vizualizacijskih tehnik, ki sem jih že omenil v teoretičnem uvodu tega dela. Osnovni gradniki so naslednji:

- a.) Datoteka 1 in 2: v datoteko 1 sem naložil .txt datoteko, ki vsebuje podatke o številu in po UDK –ju razvrščenih poizvedb, medtem ko sem v datoteko 2 naložil .txt datoteko, ki je izid čiščenja podatkov oziroma poizvedb in vsebuje podatke o rangih, številu, **K**, besedah ter razvrščenih besed po UDK.

- b.) V podatkovni tabeli je možno imeti na vpogled obe zbirki podatkov (*gl. povezavo med datoteko 1 in 2 in podatkovno tabelo, ki je povezana z obema datotekama*).
- c.) Datoteka 1 se povezuje s statistiko 1 (*statistika atributov*), distribucijo 1, Sievovem diagramu in Scatter Plotom. Datoteka 2 se povezuje s statistiko 2, distribucijo 2, Scatter Plotom 2 in matriko. V nadaljevanju bom po vrstnem redu opisal lastnosti teh vizualizacijskih tehnik.
1. Statistika 1: prikazuje statistične vrednosti za attribute v datoteki 1. Atribut »Število« nam pokaže največje in najmanjše število poizvedb v tej podatkovni množici, povprečno vrednost in mediano. Povrhu tega posreduje število vseh zajetih in različnih vrednosti (8235, 17). Ko osvetlimo UDK dobimo na vpogled stolpčni diagram, ki pokaže v številkah pogostost pojavljanja posameznih UDK področij neupoštevajoče frekvenc.
  2. Distribucija 1: na voljo imamo dve opciji to splošno (*General*) in nastavitve (*Settings*). Pod »Splošno« je možno videti porazdelitev znotraj podameznih UDK področij, krivuljo verjetnosti in zaupanja, če smo predhodno ta dva parametra nastavili.
  3. Scatter Plot 1: primerja UDK področja (X os) in število (Y os). Pod opcijo nastavitve sem nastavljal »Pokaži vse attribute«.
  4. Sieve diagram: v celotni sliki v preglednični obliki prikaže velikost posameznih UDK področij. Pod opcijo »Pogoj (*Condition*)« nastavim atribut tj. UDK in določim njegovo vrednost od 0 do 9. Na takšen način lahko imam podroben vpogled v številčnosti poizvedb po posameznih UDK področjih.
  5. Statistika 2: dobim vpogled v naslednje lastnosti podatkov kot so rang (*najvišji rang = 1, najnižji rang = 1144; število različnih vrednosti je 80; skupno število vrednosti = 1155; mediana = 473 in 267 in povprečje = 370,81 +/- 180,90*), število ali frekvenca (*najvišja frekvenca = 91, najnižja frekvenca = 2; mediana od 2 do 4; povprečje = 4,12 +/- 5,42; z 32 različnimi vrednostmi in celokupno vrednostjo, ki je enaka 1155*), **K** (*najvišja vrednost K = 68,65, najnižja vrednost K = 2,09; mediana = 4,16; povprečje = 4,67 +/- 4,75, z 80 različnimi vrednostmi in celokupno vrednostjo = 1155*), besede (*na vpogled dobim vse besede, ki so dosegle ustrezen K*) in stolpčni diagram, ki v številkah prikazuje pogostost pojavljanja posameznih UDK področij neupoštevajoče frekvenc.
  6. Distribucija 2: podobno kot pri »Distribucija 1«, le s to razliko, da dobim še vpogled v porazdelitev rangov, besed, frekvenc in K. Pod opcijo nastavitve sem prav tako nastavljal krivuljo verjetnosti in zaupanja.
  7. Scatter Plot 2: med sabo je možno primerjati vse kazalce, ki sem jih maloprej navedel. Prav tako sem tudi v tem primeru nastavljal pogoj »Prikaži vse attribute«.

8. Scatter Plotova matrika: na eni strani lahko imamo vpogled v vse primerjave tj. med rangom in K, frekvenco in rangom, UDK-jem in rangom, UDK-jem in K-jem, frekvenco in K ipd. Vizualizacijska tehnika je v tem primeru manj pregledna, in zato v tem primeru bolj kot ne služi v smeri pridobivanja občutka v zvezi z različnimi primerjavami kazalcev.

Preden bom načel četrto poglavje bom še pripravil podatke za ugotavljanje asociacijskih povezav in hierarhičnih grozdov v obliki slikovnega in besednega opisa.

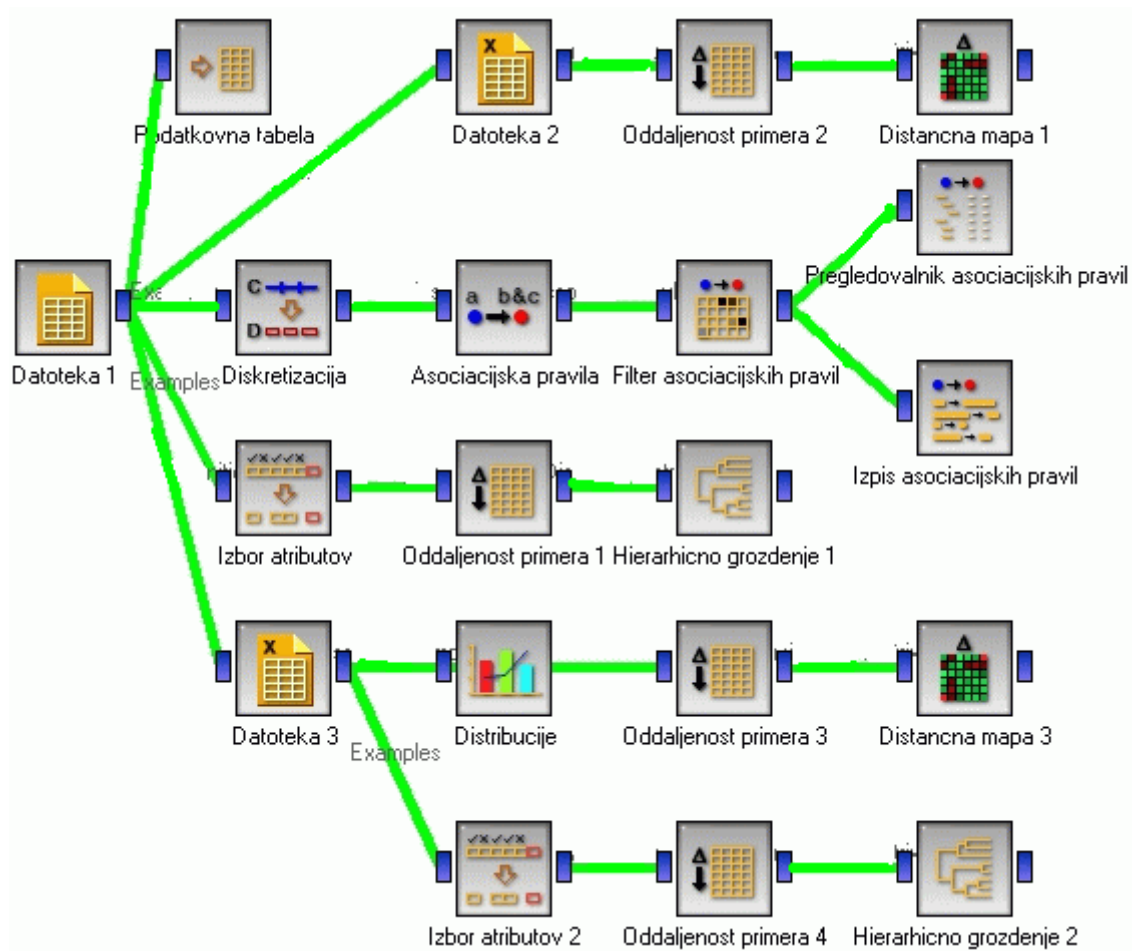
Priprava podatkov za ugotavljanje asociacijskih povezav se sestoji iz dveh skoraj vzporednih delov. Z ozirom na asociacijska pravila številčnih podatkov bom izbral najbolj primerno asociacijsko pravilo glede na dejavnika **K** in frekvence (*na naslednji strani sledi opis*), nakar bom izbral in pripravil podatke za ugotavljanje asociacijskih povezav najbolj pomembnih ključnih besed s pomočjo programskega orodja DM II – CBA.<sup>31</sup> To programsko orodje je posebej izdelano za ugotavljanje asociacijskih povezav besed, kar je tudi eden od temeljnih razlogov, da sem se odločil zanj, kajti delo s tem programom je enostavno in pregledno. Podatke je potrebno pripraviti v .tra (*Transaction Association Format*) datotečno obliko in jo ustrezno poimenovati. V nadaljevanju naložimo ustrezno .tra datoteko (*textmine37.tra*) in jo najprej z opcijo »Pretvornik, čistilec podatkov (*Data Convertor, Cleaner*)« očistimo, nakar s pomočjo ukaza »Začni OZVP (*Start Mining*)« izdelamo asociacijska pravila. Pri tem je predhodno potrebno določiti določene omejitve tako glede dejavnika podpore (*določimo minimalno podporo 5 %*), zaupanja (*določimo minimalno zaupanje 50 %*), omejitve števila pravil na največ 10000 kot tudi določiti dolžino izdelanega pravila (*Conds = 4*). Asociacijska pravila si je možno shraniti (*npr. v .doc obliki*) in si jih ogledati. Kot drugo možnost programsko orodje CBA 1.0 ponudi tudi možnost drevesnega pogleda, ki je zelo nazoren. Asociacijska pravila z ozirom na besedno analizo poizvedb (*izbral sem 34 najbolj udarnih ključnih besed in ustrezne vsebinske povezave in tako izdelal primerno .tra datoteko*) bom v poznejšem poglavju pripravil v obliki preglednice. Vseh asociacijskih pravil je bilo 80, ki sem jih tudi razvrstil po UDK. Prav na podlagi po UDK razvrščenih asociacijskih pravil bo možno ugotoviti, na katerem področju človekovega znanja so bila najmočnejša asociacijska pravila. Prav ta izid bom potem primerjal z dobljenim izidom pri ugotavljanju asociacijskih povezav številčnih podatkov z ozirom na pomembne ključne besede.

V drugem (*vzporednem*) delu priprave podatkov za asociacijskih pravila se bom kot že malo prej omenjeno ukvarjal s številčnimi podatki, ki so bili izid izvedene besedne analize

---

<sup>31</sup> Programsko orodje CBA različica 1 so izdelali na National University of Singapore.

s pomočjo programskega orodja AntConc. Prav slikovni in besedni opis tega drugega dela bo možno videti na naslednjih straneh.



### 3.9.2 Slikovni prikaz 12: Priprava podatkov za ugotavljanje asociacijskih povezav in hierarhičnih grozdov

Slika 12 prikazuje drugi del tretje stopnje priprave podatkov, ki pomeni nadaljevanje prvega dela in je s tem posledično prav tako izhodišče za kasnejšo OZVP. V tej drugi osnovni shemi so zajeti strukturirani podatki, ki so izid prve in druge stopnje priprave podatkov in vizualizacijskih tehnik, ki sem jih že omenil v teoretičnem uvodu tega dela. Osnovni gradniki so naslednji:

1. Datoteka 1: antconcisto0306.txt, 1155 primerov, štirje atributi ( $r$ ,  $N$ ,  $K$ ,  $B$ ) in diskretni razred z 10 različnimi vrednostmi. Datoteka 1 je povezana z diskretizacijo, v nadaljevanju se slednje povezuje z asociacijskimi pravili, s filtrom za asociacijska pravila in ta je povezan s pregledovalnikom za asociacijska pravila ter izpisovalnikom za asociacijska pravila. Na slikovnem prikazu 12 je možno videti, da se (na sliki je to prikazano vzporedno), da se datoteka 1 povezuje z »Izbor atributov 1«, ta je povezan z gradnikom »Oddaljenost primera 1 in Hierarhično grozdenje 1«. Datoteka 1 je tudi povezana z

datoteko 2 in datoteko 3. Preden bom opisal slednji datoteki in z njima povezane gradnike, bo najprej potrebno na kratko predstaviti poprej naštetih gradnike.

a.) Diskretizacija: metoda enakih širokih razmikov intervalov ( $r$ ,  $N$ ,  $K$ ), število intervalov je pet. Nastavitve se lahko pozneje spremenijo.

b.) Asociacijska pravila: izločanje vrednosti, ki dosegajo samo do 10 % podpore (*support*) in 50 % zaupanja (*confidence*), največje možno število pravil je 10000, nakar se s pomočjo ukaza »Izgradi pravila (*Build rules*)« izdelajo pravila. Algoritem bom določil v naslednjem poglavju, morda se bom odločil za »Vpeljevanje klasifikacijskih pravil (*Induce Classification Rules*)?«

c.) Filter asociacijskih pravil: podpora od 36 – 96 %, zaupanje od 37 – 100 %, število pravil je 22.

d.) Pregledovalnik asociacijskih pravil:  $D_K \leq 15,404$ ;  $D_N \leq 19,80$ ;  $D_r = [458,20, 686,80]$ ;  $UDK = 9$ .

e.) Izpisovalnik asociacijskih pravil:

Supp	Conf	Rule
0.377	0.386	$D_N \leq 19.80 \rightarrow UDK=9$
0.364	0.375	$D_K \leq 15.404 \rightarrow UDK=9$
0.361	0.376	$D_N \leq 19.80 \wedge D_K \leq 15.404 \rightarrow UDK=9$
0.361	0.372	$D_K \leq 15.404 \rightarrow D_N \leq 19.80 \wedge UDK=9$
0.361	0.370	$D_N \leq 19.80 \rightarrow D_K \leq 15.404 \wedge UDK=9$
0.585	0.600	$D_N \leq 19.80 \rightarrow D_r = [458.20, 686.80)$
0.585	0.603	$D_K \leq 15.404 \rightarrow D_r = [458.20, 686.80)$
0.585	0.610	$D_N \leq 19.80 \wedge D_K \leq 15.404 \rightarrow D_r = [458.20, 686.80)$
0.585	0.603	$D_K \leq 15.404 \rightarrow D_r = [458.20, 686.80) \wedge D_N \leq 19.80$
0.585	0.600	$D_N \leq 19.80 \rightarrow D_r = [458.20, 686.80) \wedge D_K \leq 15.404$
0.364	0.935	$UDK=9 \rightarrow D_K \leq 15.404$
0.361	0.929	$UDK=9 \rightarrow D_N \leq 19.80 \wedge D_K \leq 15.404$
0.361	0.959	$D_N \leq 19.80 \wedge UDK=9 \rightarrow D_K \leq 15.404$
0.377	0.969	$UDK=9 \rightarrow D_N \leq 19.80$
0.361	0.993	$D_K \leq 15.404 \wedge UDK=9 \rightarrow D_N \leq 19.80$
0.585	1.000	$D_r = [458.20, 686.80) \rightarrow D_N \leq 19.80$
0.585	1.000	$D_r = [458.20, 686.80) \rightarrow D_K \leq 15.404$
0.585	1.000	$D_r = [458.20, 686.80) \wedge D_K \leq 15.404 \rightarrow D_N \leq 19.80$
0.585	1.000	$D_r = [458.20, 686.80) \wedge D_N \leq 19.80 \rightarrow D_K \leq 15.404$

0.585 1.000  $D_r=[458.20, 686.80) \rightarrow D_N=<19.80$   $D_K=<15.404$

0.960 0.989  $D_K=<15.404 \rightarrow D_N=<19.80$

0.960 0.984  $D_N=<19.80 \rightarrow D_K=<15.404$

f.) Izbor atributov 1: atribut K, atributi na voljo r in N, razred UDK, meta razred B.

g.) Oddaljenost primera 1: distančna metrika (*Distance metrics*) evklidski algoritem, K.

h.) Hierarhično grozdenje 1: dendrogram z UDK anotacijo, povprečna povezanost (*Average Linkage*). Uporabil bom tudi podatke o številu obiskov po posameznih UDK področjih ( *vključil sem spletometrične podatke od 2003 do 2006*), kar sem izmeril s spletnim programskim orodjem Nedstat. Število obiskov po posameznih 0,1,2,3,5,6,7,8,9 UDK področjih: 5820, 7479, 9796, 10910, 10237, 8174, 11991, 13830, 15573, medtem ko je bilo število obiskov na glavni (*indeksni*) strani 16115.

2.) Datoteka 2: asomoc.txt, 9 primerov, 9 atributov, klasifikacija – diskretni razred s tremi vrednostmi.

a.) Oddaljenost primera 2: Evklidski algoritem.

b.) Distančna mapa 1: uporabljen barvni spekter je zelen (*zelo močna povezanost*) – črn (*močna povezanost*) – rdeč (*šibkejša povezanost*).

3.) Datoteka 3: asprav0306.txt, 22 primerov, 6 atributov, diskretni razred z 22 različnimi vrednostmi. Datoteko 3 sem dobil na podlagi shranjevanja gradnika »Izpis asociacijskih pravil« v .txt datoteko.

a.) Distribucije: vizualni prikaz porazdelitve asociacijskih pravil.

b.) Oddaljenost primera 3: relief, podpora.

c.) Distančna mapa 3: barvni spekter je zelen (*zelo močna povezanost*) – črn (*močna povezanost*) – rdeč (*šibkejša povezanost*).

d.) Izbor atributov 2: atributa podpora, zaupanje in razred pravilo.

e.) Oddaljenost primera 4: relief, podpora.

f.) Hierarhično grozdenje 2: dendrogram z anotacijo »Podpora« in uporabljena metoda povprečne povezanosti. Tretje poglavje načrtovanja in priprave podatkov je tako zaključeno. V naslednjem poglavju bom izbral najbolj izrazite tehnike, izgradil ustrezen glavni model in podmodele za OZVP, natančneje bom določil parametre in povezal bom med sabo tiste podatke, ki mi bodo dajali najbolj zanimive oziroma za interpretacijo uporabne vzorce, ki jih bo možno učinkovito ovrednotiti. S pomočjo tovrstnega pristopa si bom pri izdelavi kvalitativnega modela adaptivne DK precej olajšal delo.

## 4 Odkrivanje zakonitosti v podatkih v okviru modela

Pričujoče poglavje je namenjeno odkrivanju zanimivih vzorcev v podatkih **oz.** OZVP, ki sem jih v prejšnjem poglavju okvirno pripravil, kar pomeni, da bo lahko na tej stopnji že zaradi posebnih pristopov, novih odkritij ipd. prišlo do določenih sprememb. Ob tem naj bo še zapisano, da si bom pri iskanju oziroma okrivanju zanimivih vzorcev v podatkih pomagal še s posebnimi razvitimi slikovnimi modeli, tako da si bodo lahko tudi bralci lažje predstavljali to razmerje moči med različnimi zanimanji spletnih obiskovalcev kot tudi, da si bodo morda tudi zbliževali njihove miselne svetove. Preden to snov predstavim, bi želel še predstaviti določena pomembna dejstva, do katerih sem se dokopal s pomočjo dolgoletnega opazovanja, zbiranja, klasificiranja poizvedb, uvoda, pisanja drugega poglavja, načrtovanja in priprave podatkov.

- 1.) Spletni obiskovalci so v 90 – 95 % vseh primerih na UDK leksikon prispeli s pomočjo dveh zunanjih iskalnikov (*med njima je še zlasti zelo pomemben Google, na drugem mestu je bil iskalnik Najdi.si*). Iskalnika predstavljata glavna vhoda (*Input*) v ta omenjeni spletni izvedbeni sistem.
- 2.) Spletni obiskovalci so pri poizvedovanju na teh zunanjih iskalnikih uporabili določene ključne besede in/ali gesla, ki jih je vseboval tudi UDK leksikon. Izidi tovrstnih poizvedb na spletnih iskalnikih so bili mnogokrat uvrščeni med prvimi desetimi zadetki.
- 3.) Spletni obiskovalci so pretežno (*izjeme so morda t.i. profesionalni poizvedovalci, ki poznajo različne ukaze in zgradbe spletnih dokumentov na svetovnem spletu*) na teh zunanjih spletnih iskalnikih poizvedovali na podlagi lastne miselne hierarhije, ki pa so jo v nadaljevanju (*v primerih, da ni bilo zadetka ali pa zadetki niso bili dovolj izčrpni*) prilagodili vsebinski hierarhiji spletnih dokumentov, ki so bili navedeni kot relevantni zadetki.
- 4.) Ob prihodu spletnih obiskovalcev na UDK leksikon so ti uporabili notranji iskalnik v naslednjih primerih:
  - zaradi slabše orientacije oziroma zaradi težave navigacije na prispeli spletni strani.
  - spletni obiskovalci so želeli dobiti še več koristnih informacij o določeni temi, ki jih je zanimala.

- spletni obiskovalci so želeli ugotoviti kako notranji iskalnik na UDK leksikonu deluje oziroma kakšne so njegove zmogljivosti. Ta predpostavka bi v določeni meri lahko obrazložila, zakaj so nekateri redki spletni obiskovalci poizvedovali z nekoliko manj dogovorjenimi poizvedbami, ki sem jih kot klasifikator potem razvrstil pod področje UDK 4.
- spletni obiskovalci so se želeli izogniti brskanju po sorazmernih številnih in obsežnih podstraneh leksikona, tako da so notranji iskalnik uporabili kot orodje, ki jim prihrani čas in energijo.
- spletni obiskovalci, ki so iskali po UDK vrstlicah (*teh je bilo sorazmerno malo – okoli 400 poizvedb*) so večinoma uporabili notranji iskalnik.
- po moji oceni je največ spletnih obiskovalcev uporabilo notranji iskalnik, ker so se želeli dokopati do kratkih in jedrnatih opredelitev glede nekaterih pojmov oziroma tujk, ki so bile večinoma strokovne narave, kajti UDK leksikon je v bistvu za tovrstno razreševanje tujih (*strokovnih*) pojmov najbolj primeren (*notranji iskalnik indeksira tudi vsebine na spletnem slovarju Slovarju.com, Šola je nora in izobraževalni zbirki Svarog*). To pomeni, da so nekateri spletni obiskovalci uporabili notranji iskalnik, ker so iskali snov za pisanje seminarske naloge (*tovrstne poizvedbe so bile v bistvu posledica šolskega programa*).
- vsebinski potencial je v največji meri prilagojen srednješolski populaciji, zato v bistvu tudi ne preseneča (*hipotetično*) dejstvo, da je bilo največ spletnih obiskovalcev na UDK leksikonu mlajšega datuma tj. od 15 do 20 let. Po drugi strani pa je bilo v opazovanem obdobju izvedeno tudi veliko število kompleksnih ali zapletenih poizvedb, ki bi jih prej pripisal spletnim obiskovalcem, ki so nekoliko starejši.

5.) Poizvedbe na notranjem iskalniku v bistvu predstavljajo nekakšne izhodne enote (*Output*) tega spletnega sistema. Spletni obiskovalci so uporabili notranji iskalnik kot orodje za razreševanje lastnega informacijskega problema in so ga uporabili še zlasti zaradi visoke stopnje pričakovanja, ki so ga gojili do njega (*poenostaviti si delo oziroma zmanjševanje napora in prihranka energije*).

6.) Organizacija spletnih dokumentov je bila za nekatere spletne obiskovalce manj oprijemljiva in so se morda zato v nekaterih primerih odločili za uporabo notranjega iskalnika. Sistem UDK je v bistvu dogovorjena umetna tvorba, kar tudi pomeni, da se je najprej potrebno nekatere znake naučiti, da lahko uporabimo ta znanja za lažje iskanje podatkov/informacij.

7.) Skratka spletni sistem (*oziroma izhodne enote - poizvedbe*), ki ga v bistvu v tem delu proučujem ima svoj vsebinski in organizacijski potencial, ki je

nekakšna manjša prvina nekaterih večjih sistemov (*vhodne enote – zunanji svetovni in domači iskalniki kot npr. Google in Najdi.si*). Pri proučevanju sistema sem si tako pomagal tudi s klasifikacijskim sistemom, s katerim razvrščam t.i. izhodne enote oziroma poizvedbe, ki so jih izvedli spletni obiskovalci.

- 8.) S poprej navedeno točko je potrebno poudariti morda najbolj pomembno dejstvo. Spletni izvedbeni (*pod-*) sistem UDK leksikon sem v letih od septembra 2001 do septembra 2002 izdelal na podlagi informacijski potreb srednješolskih dijakov in profesorjev kot tudi od maja 2002 do decembra 2002 na podlagi poizvedb, ki so jih izvedli spletni obiskovalci na omenjenem leksikonu. Sama organiziranost podatkov/informacij na spletni izvedbi je možno tako delno poistovetiti s postavitvijo knjižničnega gradiva v srednješolski knjižnici, kjer sem kot bibliotekar služboval.

Prav vse navedene točke (*oziroma ozadja*) naj bi bile nekakšno uvodno vodilo pri nadaljnjem odkrivanju zanimivih vzorcev v podatkih. Poleg tega bi želel še opozoriti na moj miselni svet oziroma na svetovni nazor, ki sem ga ponovno odkril na svetovnem spletu ob branju zanimivega spletnega dokumenta.<sup>32</sup> Osnovna zamisel izhaja s predpostavke, da lahko gledamo na ves svet kot na organizem, v katerem obstajajo in nastajajo dinamična omrežja primerljiva s človeškimi možgani in svetovnim spletom. V teh omrežjih obstajajo skupine, ki se povezujejo s sorodnimi skupinami glede na globalne in posebne interese. Prav tovrstne skupine ljudi imajo nek miselni hierarhični potencial, ki je posledica genov, družinske in družbene vzgoje. S tovrstnim osnovnim programom se ljudje podajo na nadaljnjo pot razvoja. Na podlagi izkušenj in raznovrstnih situacij razvijajo lastne programe, ki pa so posledica globalnega programa. Ta globalni program pa je nastal zaradi nekega višjega programa, ki je v prvi stopnji naš planet Zemlja, ki prav tako deluje programsko pod vplivom še višjega programa, med katerimi naj omenim naše Sonce. Ljudje smo nekakšni programi, ki ustvarjajo programe. Prav programi, ki se morajo prilagajati nekemu hierarhično višjemu konceptu, pomenijo nekakšna podlaga za omrežja. Manjša omrežja se bolj ali manj prilagajajo večjemu omrežju. Spletni obiskovalci se prilagajajo kolektivnimi družbeni programi, vendar sleherni posameznik v večji ali manjši meri prispeva k raznolikosti tovrstnih programov. Ta svetovni nazor bo močno vplival na poznejšo segmentacijo (*izobraževalnih*) interesov spletnih obiskovalcev in na še kasnejšo

---

<sup>32</sup> Heylighen, F.(1996). The Social Superorganism and its global brain. Dostopno: <http://pespmc1.vub.be/HEYL.html> (2005-07-12)

ustvarjanje kvalitativnega modela adaptivne digitalne knjižnice, zato ocenjujem, da je bila kratka predstavitev tovrstne zamisli nujna že zaradi poznejšega razumevanja grozdenja ljudi v določene interesne skupine, kar bom uresničil s pomočjo UDK in iz njega razvitega naravnega modela, ki sem ga že opisal v drugem poglavju tega dela.

Zdaj v nadaljevanju bom najprej predstavil zakon moči pred in po čiščenju podatkov.

#### 4.1 Preglednica 4: Zakon moči pred in po čiščenju podatkov

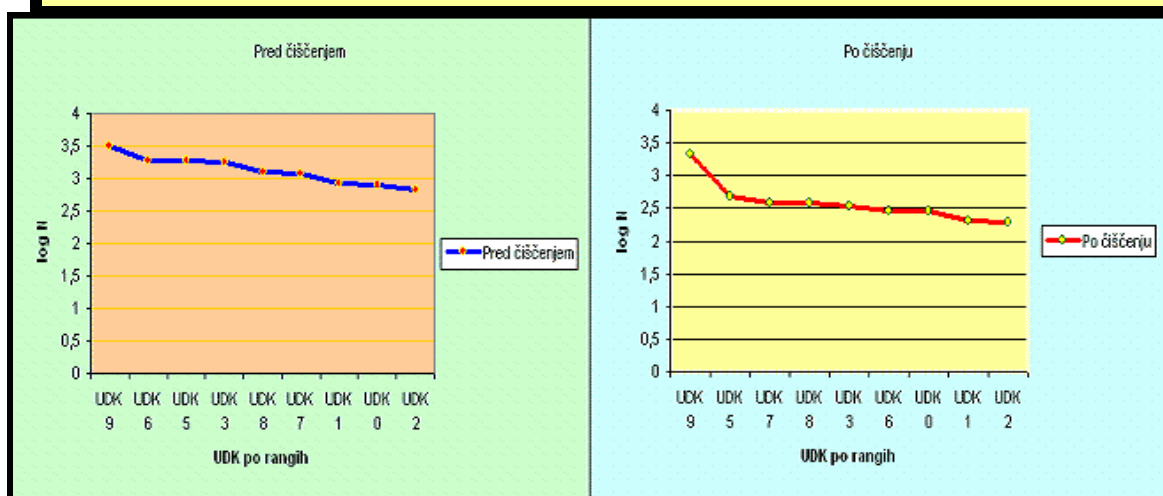
	UDK 0	UDK 1	UDK 2	UDK 3	UDK 5	UDK 6	UDK 7	UDK 8	UDK 9
<b>N P</b>	809	846	660	1828	1863	1917	1155	1256	3127
<b>log NP</b>	2,9079	2,9283	2,8195	3,2619	3,2702	3,2826	3,0625	3,0989	3,4951
<b>R P</b>	8	7	9	4	3	2	6	5	1
<b>% P</b>	5,95	6,21	4,85	13,43	13,68	14,08	8,48	9,23	22,97
<b>N B</b>	284	205	191	351	491	293	389	374	2186
<b>log NB</b>	2,4533	2,3117	2,2810	2,5453	2,6910	2,4668	2,5899	2,5728	3,3396
<b>R B</b>	7	8	9	5	2	6	3	4	1
<b>% B</b>	5,95	4,30	4,00	7,36	10,36	6,15	8,16	7,84	45,88

Število vseh zajetih poizvedb pred čiščenjem = 13613 (100 %)

Poizvedovalni šum UDK 4 pred čiščenjem = 152 poizvedb (1,12 %)

Število besed po čiščenju na podlagi K = 4764 (100 %)

Šum UDK 4 po čiščenju = 0 besed (0,00 %)



4.1.1 Slikovni prikaz 13: Logaritmirani in po rangih razvrščeni podatki

Preglednica 4 in slika 13 prikazujeta zakon moči s pomočjo primerjave med rangiranimi UDK področij in desetiško logaritmiranimi frekvencami. Gre za prikaz stanja podatkov pred čiščenjem (*poizvedbe*) in po čiščenju (*postopek čiščenja poizvedb s pomočjo dejavnika K, pri katerem sem kot izid dobil pomembne ključne besede*). V preglednici 4 sem predstavil naslednje podatke:

a.) Pred čiščenjem

- Vodoravno: razvrščene poizvedbe po UDK področjih od 0, 1, 2, 3, 5, 6, 7, 8 in 9  
 - Navpično: NP (*število ali frekvenca poizvedb*), log NP (*logaritmirane frekvence poizvedb*), RP (*rang poizvedb*) in % P (*odstotni delež poizvedb po posameznih UDK področjih*).

b.) Po čiščenju

- Vodoravno: razvrščene ključne besede po UDK področjih od 0, 1, 2, 3, 5, 6, 7, 8 in 9  
 - Navpično: NB (*število ključnih besed*), log NB (*logaritmirane frekvence ključnih besed*), RB (*rang ključnih besed z ozirom na posamezna UDK področja*) in % B (*odstotni delež ključnih besed*). Spodnji del preglednice 4 prikazuje še podatke o celokupni frekvenci vseh poizvedb ( $N = 13613$  ali 100,00 %), podatek za poizvedovalni šum t.j. področje UDK 4 (152 poizvedb ali 1,12 %), skupna frekvenca vseh pomembnejših ključnih besed (4764 ali 100,00 %) in učinkovito čiščenje neopredeljenih poizvedb z nizko stopnjo pomena t.j. področje UDK 4 (0 besed ali 0,00 %). V preglednici 4 bom pri opisu izpostavil podatke, ki se nahajajo v obarvanih območjih (*log NP, RP, log NB in RB*), kajti tovrstni podatki so za opis slikovnega prikaza 13 ključnega pomena.

**Poizvedbe (rang; UDK; log NP): Pomembnejše ključne besede (rang; UDK; log NB)**

1. ali vrhnja Skupina

Rang 1; UDK 9; 3,4951	:	Rang 1; UDK 9; 3,3396
Rang 2; UDK 6; 3,2826	:	Rang 2; UDK 5; 2,6910
Rang 3; UDK 5; 3,2702	:	Rang 3; UDK 7; 2,5899

2. ali srednja skupina

Rang 4; UDK 3; 3,2619	:	Rang 4; UDK 8; 2,5728
Rang 5; UDK 8; 3,0989	:	Rang 5; UDK 3; 2,5453
Rang 6; UDK 7; 3,0625	:	Rang 6; UDK 6; 2,4668

3. ali spodnja skupina

Rang 7; UDK 1; 2,9283	:	Rang 7; UDK 0; 2,4533
Rang 8; UDK 0; 2,9079	:	Rang 8; UDK 1; 2,3117
Rang 9; UDK 2; 2,8195	:	Rang 9; UDK 2; 2,2810

Na podlagi razvrščenih podatkih po rangih je možno sedaj izračunati za posamezna UDK področja rangirne razdalje. Rangirne razdalje za podatke pred in po čiščenju nam povedo, za koliko rangirnih enot je določeno UDK področje porastlo ali pa pojemalo. Osnovni matematični obrazec za izračun rangirne razdalje je naslednji:  $\Delta R_{UDK} = R_{pUDK} - R_{bUDK}$   $\Delta R_{UDK}$  ... Rangirna razdalja določenega UDK področja;  $R_{pUDK}$  ... Rang določenega UDK področja z ozirom na poizvedbe;  $R_{bUDK}$  ... Rang določenega UDK področja z ozirom na moč ključnih besed.

$\Delta R_{UDK9} = R_{pUDK9} - R_{bUDK9} = 1 - 1 = 0$  Rang področja UDK 9 se tudi po čiščenju podatkov ni spremenil, tako da je področje UDK 9 ohranilo prvo mesto in se tako uvršča v prvo ali vrhno skupino visokega vsebinskega pomena. Prav tako je razlika med logaritmiranima frekvencama dokaj nizka, saj znaša 0,1555.

$\Delta R_{UDK6} = R_{pUDK6} - R_{bUDK6} = 2 - 6 = -4$  Rang področja UDK 6 se je po čiščenju podatkov precej spremenil, tako da je področje UDK 6 z drugega mesta padlo na šesto mesto. Po čiščenju podatkov se je vsebinski pomen tovrstnih podatkov precej znižal, tako da se področje UDK 6 po čiščenju podatkov uvršča v drugo ali srednjo skupino. Prav tako je razlika med logaritmiranima frekvencama precej višja, saj znaša 0,8158. Prav to nas posledično napelje na ogromno razliko med logaritmiranima frekvencama med UDK 9 in UDK 6 po čiščenju podatkov, ki znaša 0,8728! Prav to razliko je možno še posebej nazorno opaziti na slikovnem prikazu 13.

$\Delta R_{UDK5} = R_{pUDK5} - R_{bUDK5} = 3 - 2 = 1$  Področje UDK 5 je porastlo za eno mesto navzgor. Razlika med logaritmiranima fekvencama znotraj tega področja znaša 0,5791, kar je dokaj visoka vrednost. Če zdaj med sabo primerjam logaritmirane frekvence po čiščenju podatkov med UDK 9 in UDK 5 dobim kot izid vrednost 0,6485! Tudi tovrstno razliko je možno lepo videti na sliki 13. Pred čiščenjem podatkov ta razlika ni bila tako očitna, saj je znašala vrednost samo 0,2249! Kakovostno gledano se je področje UDK 5 obdržalo v prvi ali vrhni skupini.

$\Delta R_{UDK3} = R_{pUDK3} - R_{bUDK3} = 4 - 5 = -1$  Področje UDK 3 je padlo za eno rangirno mesto. Razlika med področjema UDK 9 in UDK 3 je prav tako zelo očitna (*gl. slika 13*). Kakovostno gledano se je področje UDK 3 obdržalo v srednji ali drugi skupini.

$\Delta R_{UDK8} = R_{pUDK8} - R_{bUDK8} = 5 - 4 = 1$  Področje UDK 8 se je dvignilo za eno rangirno enoto in prav tako ohranilo mesto v drugi ali srednji skupini.

$\Delta R_{UDK7} = R_{pUDK7} - R_{bUDK7} = 6 - 3 = 3$  Področje UDK 7 se je dvignilo za tri rangirna mesta, tako da se je kakovostno gledano povzpela v prvo ali vrhnjo skupino. Na slikovnem prikazu 13 je možno ta dvig zelo lepo videti.

$\Delta R_{UDK1} = R_{pUDK1} - R_{bUDK1} = 7 - 8 = -1$  Področje UDK 1 je padlo za eno rangirno mesto in tako ohranilo mesto v tretji ali spodnji skupini.

$\Delta R_{UDK0} = R_{pUDK0} - R_{bUDK0} = 8 - 7 = 1$  Področje UDK 0 se je povzpelo za eno rangirno enoto navzgor, vendar kljub temu po vsebinski moči ostalo v tretji skupini.

$\Delta R_{UDK2} = R_{pUDK2} - R_{bUDK2} = 9 - 9 = 0$  Področje UDK 2 je hranilo tako tudi po čiščenju podatkov, saj je ohranilo rang 9 kot tudi mesto v tretji ali spodnji skupini.

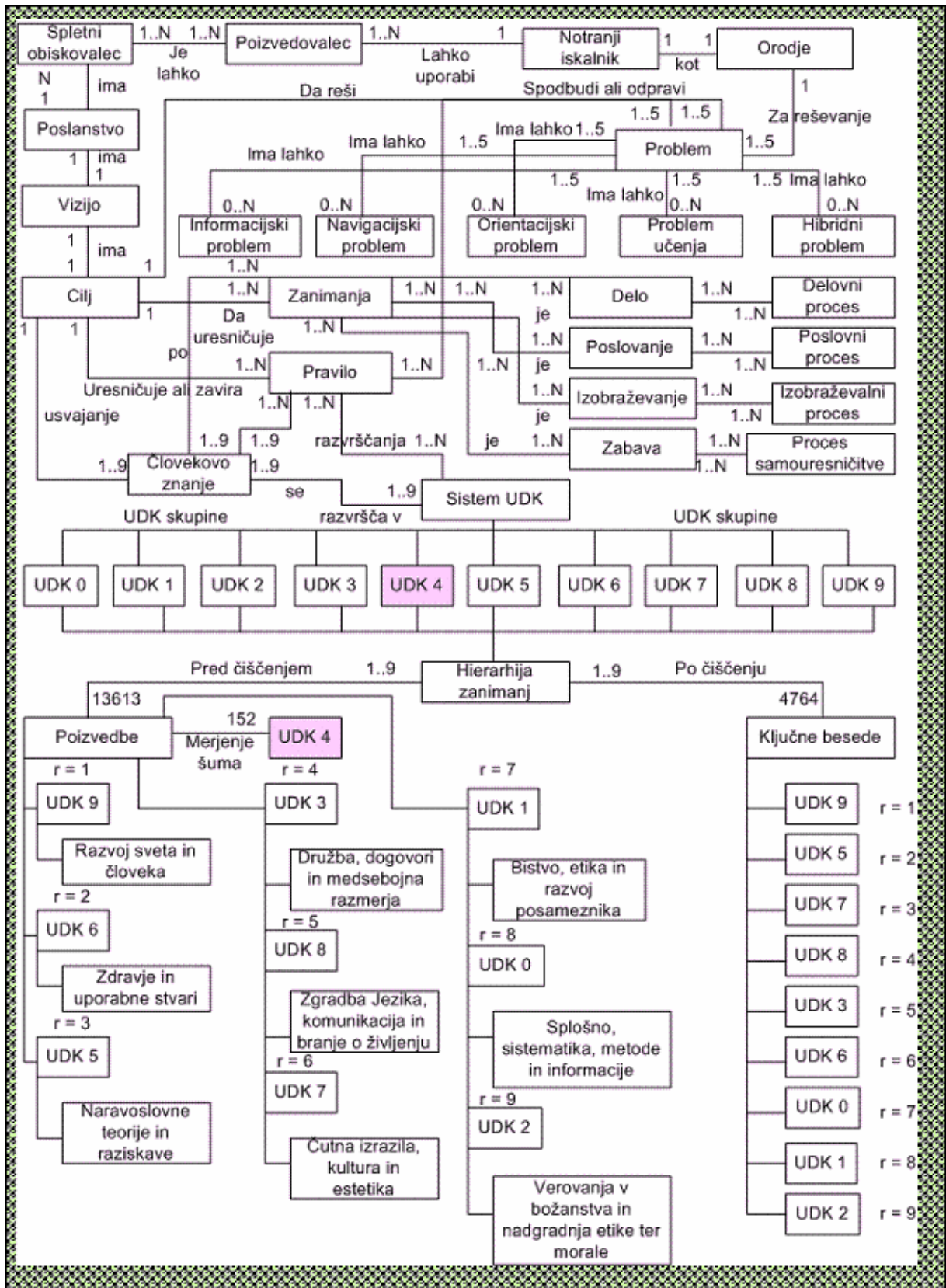
Sklep 1: razlike med logaritmiranimi frekvencami po čiščenju podatkov dosegajo največji odmik takrat, kadar med sabo primerjam področje UDK 9 z drugimi UDK področij.

Sklep 2: ko primerjam med sabo razlike logaritmiranih frekvenc po čiščenju podatkov med različnimi UDK področij razen UDK 9 prihajam do spoznanja, da so dokaj nizke razlike.

Sklep 3: S kakovostnega vidika je možno ugotoviti, da sta po primerjavi pred in po čiščenju podatkov zgolj področji UDK 7 in UDK 6 spremenili skupini, tako je področje UDK 7 prispelo v vrhnjo, medtem ko je področje UDK 6 padlo v srednjo skupino (*gl. velike rangirne razdalje*). Z vidika pomembnosti ali moč ključne besede (K) lahko ocenjujem, da so podatki po postopku čiščenja s področja UDK 7 pridobili na pomenu, medtem ko so podatki s področja UDK 6 po postopku čiščenja UDK 6 izgubili na pomenu.

Sklep 4: če je možno za levi del slike 13 trditi, da obstaja v zvezi s poizvedovanjem na UDK leksikonu neko stabilno stanje, tega ni možno trditi za desni del slike 13, kajti pomen ključnih besed, ki spadajo v področje UDK 9 je izjemno visok nasproti drugim ključnim besedam z drugih UDK področij. Postopek čiščenja podatkov je tako izpostavil pomembne ključne besede in izločil mnogo manj pomembnih. Prav tako je postopek čiščenja podatkov izločil tudi poizvedbe (*poizvedovalni šum*), ki sem jih pred čiščenjem podatkov razvrstil pod UDK 4.

Sklep 5: ocenjujem, da je možno na podlagi dobljenih izidov vzpostaviti globalni miselni model poizvedovalcev oziroma spletnih obiskovalcev, kajti na podlagi področnih (*izobraževalnih*) interesov, ki so jih izrazili s poizvedbami na notranjem iskalniku je le-te že možno razvrstiti v določene interesne UDK skupine. Prav tako bi bilo že možno, še zlasti pridobljene pomembne ključne besede med sabo povezovati in tako ustvariti določena asociacijska pravila, kar bom okvirno prikazal na naslednjih straneh.



**4.2 Slikovni prikaz 14: Okvirna miselna hierarhija interesnih področij spletnih obiskovalcev – poizvedovalcev na UDK leksikonu**

Slika 14 prikazuje okvirno miselno hierarhijo spletnih obiskovalcev z ozirom na interesna (izobraževalna) področja, ki sem jo ustvaril na podlagi izidov spletometrične analize s pomočjo metode zakona moči. Vrhnji del slike 14 predstavlja glavne težnje spletnega obiskovalca oziroma uporabnika notranjega iskalnika, ki so naslednje:<sup>33</sup>

- notranji iskalnik je zanj orodje, s katerim si poskuša povečati informativnost, prihraniti energijo in čas. Skratka spletni obiskovalci, ki uporabljajo iskalnik želijo rešiti določen problem, ki sem ga razčlenil na informacijski problem (*npr. pridobivanje informacij za pisanje diplomskega dela, za poslovno odločitev*), navigacijski problem (*npr. uporabnik spletnega iskalnika ne želi po zamudnih poteh s pomočjo številnih klikov priti do zaželenih informacij, temveč si zamišlja, da je tovrstnih dejavnosti čim manj*), orientacijski problem (*npr. poizvedovalec se lokacijsko in vsebinsko gledano na spletnih straneh določene izobraževalne zbirke ne znajde najboljših in bo zato uporabil iskalnik*), problem učenja (*npr. uporabnik iskalnika želi nadgraditi svoje znanje glede uporabe iskalnika in bo zato preizkušal različne ukaze, ki mu bodo dajali čim bolj kakovostne in izčrpne izide*) in nenazadnje obstaja še hibridni problem, ki je lahko kombinacija dveh ali več poprej naštetih problemov.
- Sleherni spletni obiskovalec – poizvedovalec ima svoje poslanstvo, vizijo in cilj. Prav tako se mora držati določenih pravil, ki so tako navigacijske kot tudi vsebinske narave. Določena pravila lahko spodbudijo uresničitev cilja, medtem ko so nekatera pravila vezana na učenje, ki pa zahteva določene vložke in svoj čas. Prav tako lahko določena pravila pomagajo razrešiti probleme, medtem ko jih določena pravila v bistvu poglobljajo. Prav tako se vežejo pravila na predmet klasifikacijskega sistema, ki že v svoji osnovi vsebuje veliko različnih pravil.
- Glavni cilj spletnega obiskovalca – poizvedovalca je v glavnem vezan na razrešitev problema /-ov, na zadovoljitev lastnih informacijskih potreb oziroma na uresničitev lastnega zanimanja in nenazadnje na pridobivanje ne zgolj ustreznih podatkov/informacij, temveč na usvajanje koristnih znanj tj. na sliki 14 prikazano kot »človekovo znanje«.
- Na spletnem UDK leksikonu so podatki pretežno organizirani po sistemu UDK, ki razvršča vso človekovo znanje s pomočjo vrstilcev idr. na devet različnih skupin (*gl. sliko 13: 0, 1, 2, 3, 5, 6, 7, 8, in 9, medtem ko UDK 4 predstavlja področje neopredeljenega oziroma poizvedovalni šum*).

---

<sup>33</sup> Povezave so zaznamovane s številčnimi oznakami kot npr. 1 ali 1..N (*od ena do števila N*), ki pomenijo nivo števnosti, ki pa za osnovno razumevanje slikovnega prikaza 14 niso ključnega pomena.

- V nadaljevanju (*gl. spodnji del slike 14*) slika 14 ponazarja izide miselne interesne hierarhije spletnih obiskovalcev – poizvedovalcev z ozirom na zakon moči tj. pred čiščenjem podatkov in miselno interesno hierarhijo po čiščenju podatkov. Pri tem sem dobil naslednje izide (*UDK področja so na vrhu zaznamovana z oznako  $r = 1 - 9$* ):
  - a.) Pred čiščenjem podatkov: uporabniki notranjega spletnega iskalnika so največkrat izvedli poizvedbe, ki sem jih razvrstil pod UDK 9 tj. interesno področje razvoja sveta in človeka. V ta okvir se uvrstijo poizvedbe s področja arheologije, domoznastva, geografije, življenjepisov znamenitih ljudi, genealogija in svetovna ter domača zgodovina. Na drugo mesto so se uvrstile poizvedbe s področja UDK 6 tj. interesno področje zdravja/bolezni in uporabnih ved. V ta okvir se uvrstijo poizvedbe s področja medicine, strojništva, elektrotehike, energetike, industrije in proizvodne tehnologije, gradbeništva, prehrane idr.
  - Po pogostosti poizvedb se je na tretje mesto uvrstilo področje UDK 5 tj. interesno področje narave, okolja, naravoslovnih teorij in raziskav pojavov. V to skupino sem uvrstil poizvedbe s področja ekologije, matematike, astronomije, fizike, kemije, geologije, meteorologije, paleontologije, geodezije, antropologije, biologije idr.
  - Na četrto mesto po pogostosti poizvedb se je uvrstilo področje UDK 3 tj. interesno področje družbe, dogovorov, odnosov, pravil in medsebojnih razmerij. V to skupino sem razvrstil poizvedbe s področja sociologije, socialna komunikacija demografije, prava, politike, sociale, ekonomije, etnologije, etnografije, metrologije, pregovorov idr.
  - Na peto mesto se je uvrstilo področje UDK 8 tj. interesno področje zgradbe in pravil jezika, jezikovne komunikacije, pisanje ter branje o življenju, ki je bilo, je sedaj in bi lahko bilo ipd. V to skupino so se uvrstile poizvedbe s področja jezikoslovja in književnosti.
  - Na šesto mesto se je uvrstilo področje UDK 7 tj. interesno področje kulture, čutnih vtisov in izrazil ter estetike. Semkaj sem uvrstil poizvedbe s področja umetnosti, arhitekture, slikarstva, kiparstva, glasbe, novceslovja, filatelije, turizma, plesa, športa idr.
  - Na sedmo mesto se je uvrstilo področje UDK 1 tj. interesno področje bistva, etike, morale, značilnosti in razvoja posameznika. V to skupino sem razvrstil poizvedbe s področja filozofije, kozmogonije, alkimizma, okultnih ved, psihologije, logike, etike in morale.
  - Na osmo mesto se je uvrstilo področje UDK 0, ki pomeni pomeni interesno področje splošnega, sistematike, metodologije in informacij. V to skupino so se uvrstile poizvedbe s področja sistematike znanosti, opredelitve, pisav, pismenstva, bibliotekarstva, informatike

(s *splošnega vidika*), računalništva (*na splošno*), organizacije (*na splošno*), muzejev, galerij, slovarjev, leksikonov, enciklopedij idr.

Na zadnje deveto mesto so se uvrstile poizvedbe s področja UDK 2, ki pomeni interesno področje verovanj v božanstva in nadgradnje etike ter morale. V to skupino so se uvrstile poizvedbe s področja krščanskih in nekrščanskih verstev, mitologij, plemenskih verovanj idr.

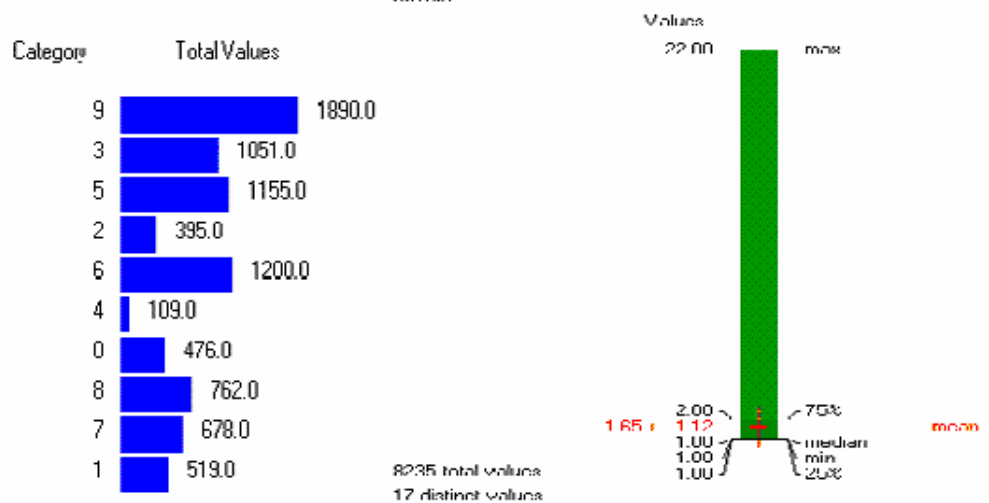
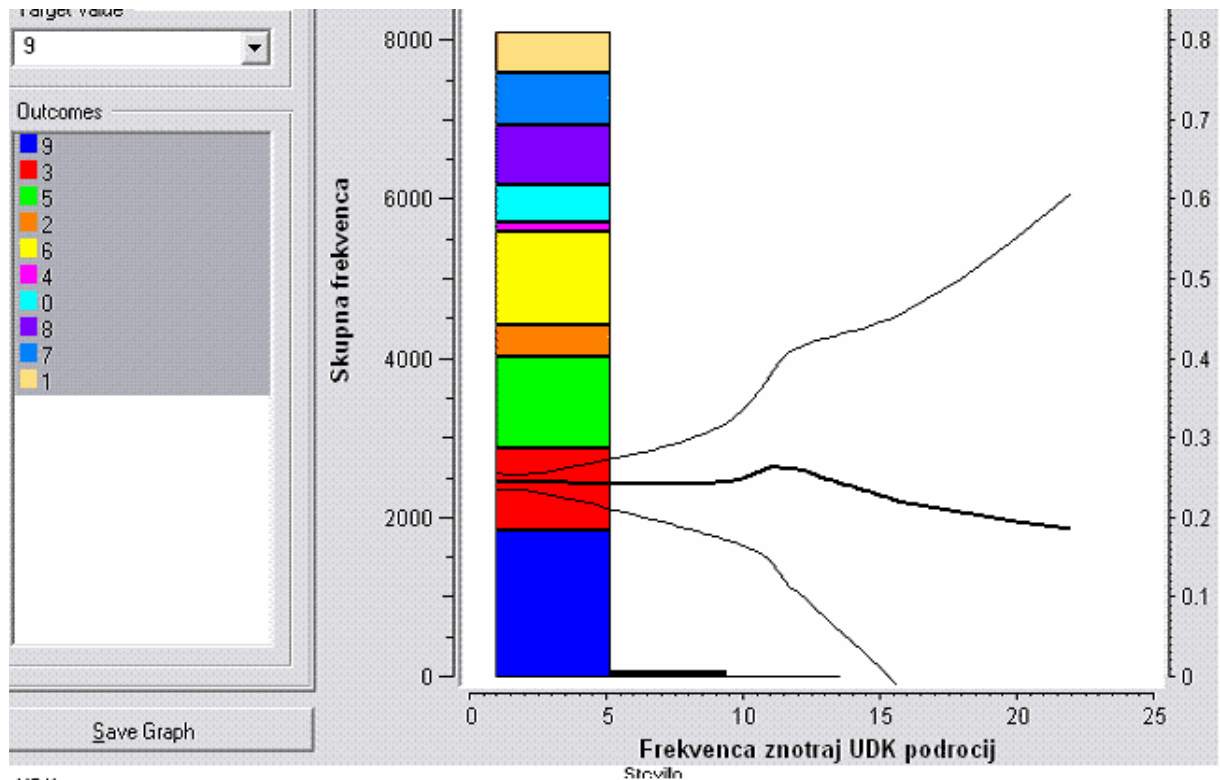
b.) Po čiščenju podatkov: rangirna lestvica z ozirom na UDK področja se je po čiščenju podatkov nekoliko spremenila zaradi besedne analize s pomočjo dejavnika **K**. Kot izid sem dobil zgolj tiste enobesedne zveze, ki so imele glede moči pomena največje vrednosti. Lestvico si lahko ogledate na spodnji desni strani slike 14. Bistveno spremembo je možno opaziti pri področju UDK 6 in UDK 7, ki sta količinsko in kakovostno gledano spremenili položaj. Pomen UDK 7 je narastel zaradi tega, ker je bilo možno nekatere poizvedbe s tega področja povezovati s poizvedbami s področja UDK 9, medtem ko pri področju UDK 6 temu ni bilo tako. Na podlagi znanih poizvedb in dobljenih izidov s pomočjo metode zakona moči lahko precej zanesljivo trdim, da so poizvedbe, ki sem jih razvrstil v področje UDK 7 v tesni povezavi s poizvedbami s področja UDK 9, še zlasti s podpodročjem UDK 9 t.j življenjepisov glasbenikov (*npr. poizvedbo »Mozartova dela« sem razvrstil pod UDK 7, medtem ko sem poizvedbo »življenje Mozarta« razvrstil pod UDK 9*).

Spodnji del slikovnega prikaza 14 pomeni nekakšno izhodišče za poznejšo segmentacijo spletnih obiskovalcev – poizvedovalcev v interesne skupine, kajti osnovna miselna interesna hierarhija poizvedovalcev je tako ugotovljena. Če zdaj na vso množico poizvedovalcev gledam kot na posamezen organizem, bom lahko s pomočjo tega miselnega prijema lažje izpeljal dejansko miselno hierarhijo in povezave z ozirom na obdelane poizvedbe, kajti poizvedbe pomenijo zunanji izraz, ki jih je ta t.i. organizem izvedel. Preden se bom lotil naslednjega podpoglavja, si naj še zastavim vprašanje, ki se navezuje na psihološke vzgibe ljudi pri ustvarjanju učinkovitega ekonomsko propagandnega programa: »Ali bi bilo možno povezati težnje poizvedovalcev na UDK leksikonu s splošno znanimi težnjami ljudi, ki lahko predstavljajo potencialni kupci določenih proizvodov ali odjemalci določenih storitev (*npr. splošno znane težnje ljudi, katere uporablja ekonomska propaganda so naslednje: prijateljstvo, ugodje, ljubezen, narod, humor, zvestoba, harmonija barv in zvokov, družinsko življenje, vplivne osebnosti, zdravje, fizična moč, mladost, naravni motivi*)?« Na to snov se bom še kasneje povrnil.

## 4.3 Izhodišče

Kot izhodišče za OZVP bom vzel slikovni prikaz 11 s strani 60, tako da bom v tem podpoglavju najprej poskusil odkrivati zanimive vzorce za podatke pred čiščenjem z že v teoretičnem uvodu omenjenimi vizualizacijskimi tehnikami. Z ozirom na dejstvo, da sem podatke opisal že pri poglavju »Priprava podatkov«, tega v tem poglavju ne bom ponovil.

## 4.4 Distribucija in statistični atributi poizvedb

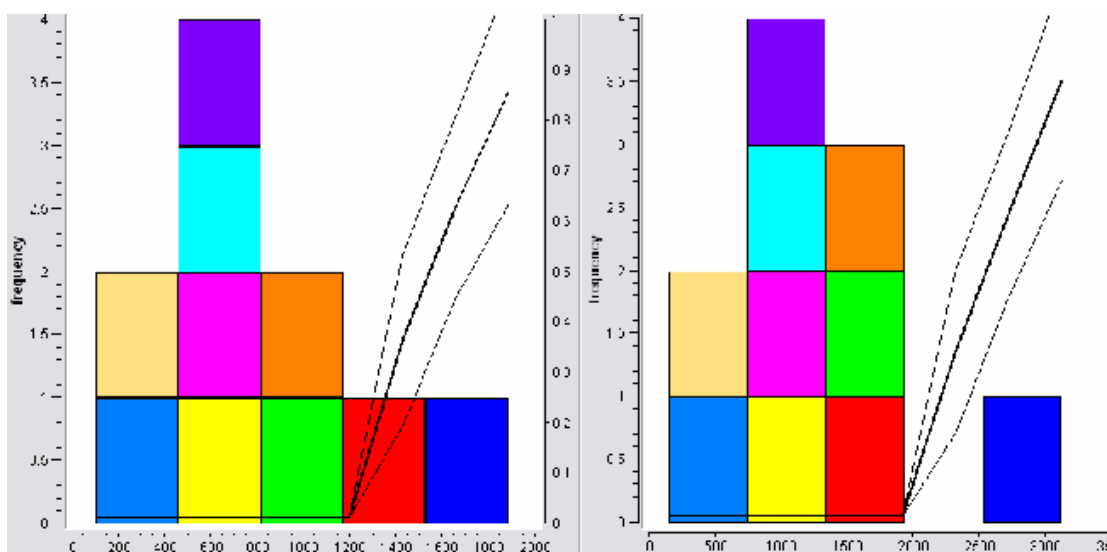


### 4.4.1 Slikovni prikaz 15: Distribucija znotraj UDK področij in statistični atributi

Slika 15 lepo ponazarja distribucijo frekvenc zgolj znotraj UDK področij in statistične attribute. Ključna ugotovitev bi bila v tem, da lahko z metodo zakona moči ponovno dokažemo stabilnost preučevanega sistema, kajti rangi posameznih UDK področij se niso spremenili. Ko primerjam med sabo frekvence znotraj UDK področij z vsemi frekvencami, dobim zelo podoben izid. Kot dokaz zapišemo naj služi naslednji preglednični in slikovni prikaz.

#### 4.5 Preglednica 5: Primerjavava med frekvencami znotraj UDK področij in vsemi frekvencami

F2+	F1	UDK	DF	log f2	log f1
3127	1890	9	1,65	3,4951	3,2765
1917	1200	6	1,6	3,2826	3,0791
1863	1155	5	1,61	3,2702	3,0625
1828	1051	3	1,58	3,2619	3,0216
1256	762	8	1,65	3,0989	2,8819
1155	678	7	1,7	3,0626	2,8312
846	519	1	1,63	2,9273	2,7151
809	476	0	1,7	2,9079	2,6776
660	395	2	1,67	2,8195	2,5965
152	109	4	1,39	2,1818	2,0374



4.5.1 Slikovni prikaz 16: Distribucija frekvenc znotraj UDK področij in UDK področij s frekvencami

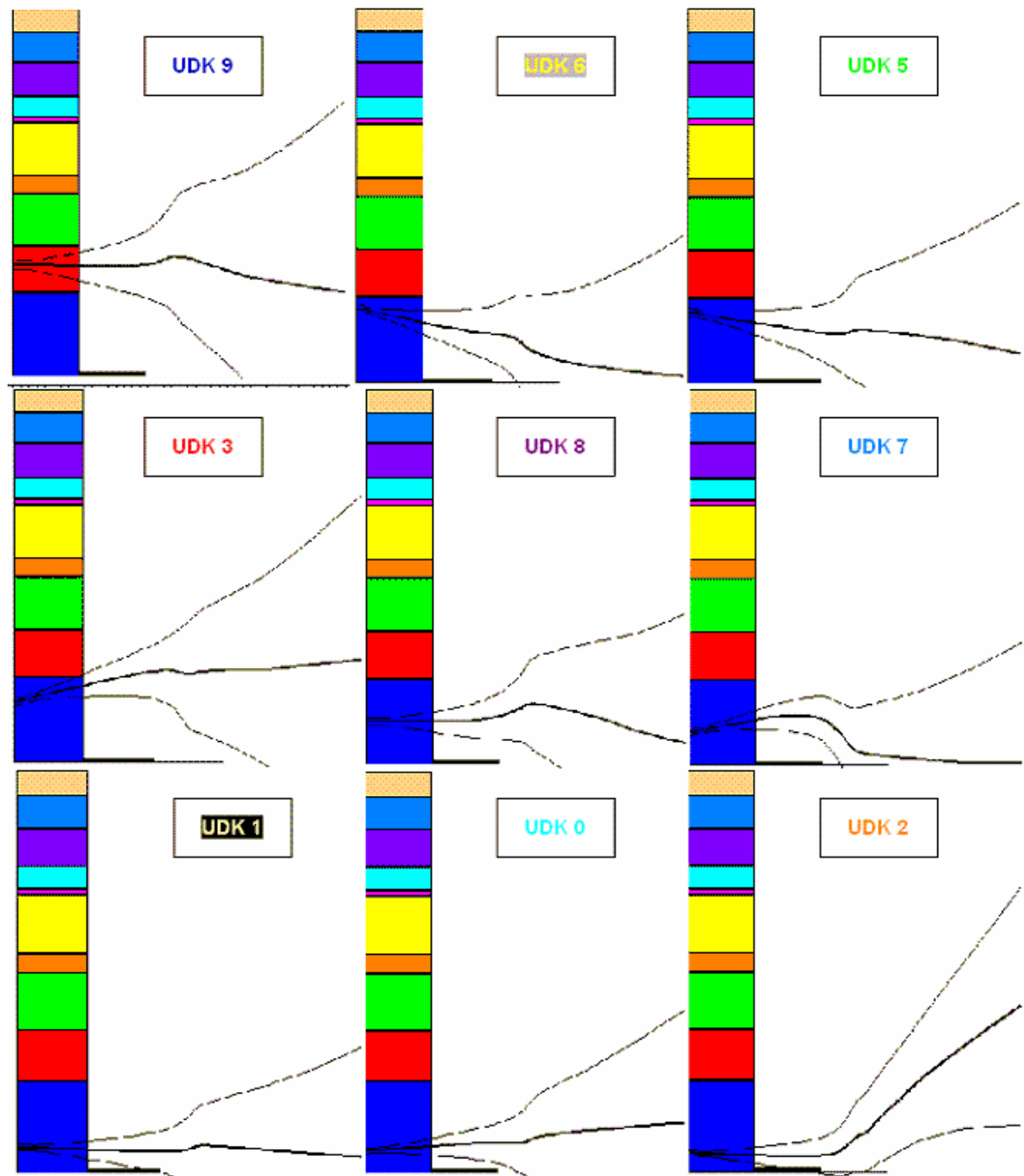
Preglednica 5 in slika 16 prikazujeta distribucijo frekvenc znotraj UDK področij (F1) in UDK področij vključno s frekvencami (F2+). UDK področja so označena z ustreznimi barvami (*temno modra = UDK 9, rdeča = UDK 6, zelena = UDK 5, oranžna = UDK 3, rumena = UDK 8, svetlo vijolična = UDK 7, svetlo modra = UDK 1, temno vijolična = UDK 0, modra = UDK 2 in modra = UDK 4*). V preglednici je možno poleg frekvenc (F1, F2+) videti razmerje med frekvencama ( $DF = F2+/F1$ ) in logaritmirane frekvence ( $\log F1, \log F2+$ ). Vsi navedeni kazalci kot tudi slikovni prikaz 16 prikažejo neko stabilno obnašanje preučevanega sistema. Glavne značilnosti določenega stabilnega sistema so naslednja:

- prevladuje red in stopnja predvidljivosti je večja
- pojav določenih relevantnih ponavljajočih se (*vedenjskih*) vzorcev je pogost
- se ne odzove na majhne spremembe in se zaradi le-teh ne zruši
- vpliv časa je manjši
- obstaja nek močan poudarek oziroma izraža neko trdno usmerjenost

Na podlagi slike 16 bi lahko sklepal, da so poizvedovalci za določeno področje poizvedovali največkrat po enkrat ali dvakrat, medtem ko so bila številnejša poizvedovanja za določen pojem z določenega področja redkejša (*velja za desno stran slike 15 tj. F2+*). Z desne strani slike 16 je možno tudi izpeljati dejstvo, da je razkorak med frekvencama (F2+) za področje UDK 9 (*prvi rang*) in UDK 5 (*drugi rang*) večji, kajti vmes ne vidimo stolpca kot je temu primer na levi strani slike. Če zdaj ponovno gledamo na vse poizvedovalce kot na celovit organizem, bi lahko trdil, da ima ta organizem stabilne usmeritve, da bi lahko bil sorazmerno predvidljiv in da ima ta organizem nek močan vsebinski poudarek.

Pri proučevanju frekvenčne porazdelitve je možno še zraven vključiti krivuljo verjetnosti in zaupanja/zanesljivosti za posamezna UDK področja. Krivulja verjetnosti za določeno UDK področje nam napove prihodnostne usmeritve, medtem ko nam krivulja zaupanja posreduje informacijo o minimalnem ali maksimalnem odmiku napovedanega. V ta namen bom ponovno uporabil podatke s podpoglavja 4.4. Pri frekvenčni porazdelitvi znotraj UDK področij brez upoštevanja frekvenc je bilo možno opaziti, da je bil pojav poizvedovalnega šuma (*UDK 4*) zelo redek in sem se zaradi tega odločil, da na naslednjem slikovnem prikazu frekvenčnih porazdelitev in krivulj verjetnosti ter zaupanj ne prikažem oba intervala za to področje. Vpliv tovrstnega šuma je z ozirom na že prej ugotovljeno stabilno stanje preučevanega sistema zanemarljive narave. Ob tem naj še poudarim, da je ena od glavnih značilnosti slehernega odprtega sistema v naravi in družbi,

da vsebuje določeno mero šuma, vendar manjša kot je stopnja šuma, tem stabilnejši je tudi sistem. Prav zadnje navedeno ponovno dokazuje, da je preučevani sistem stabilen.



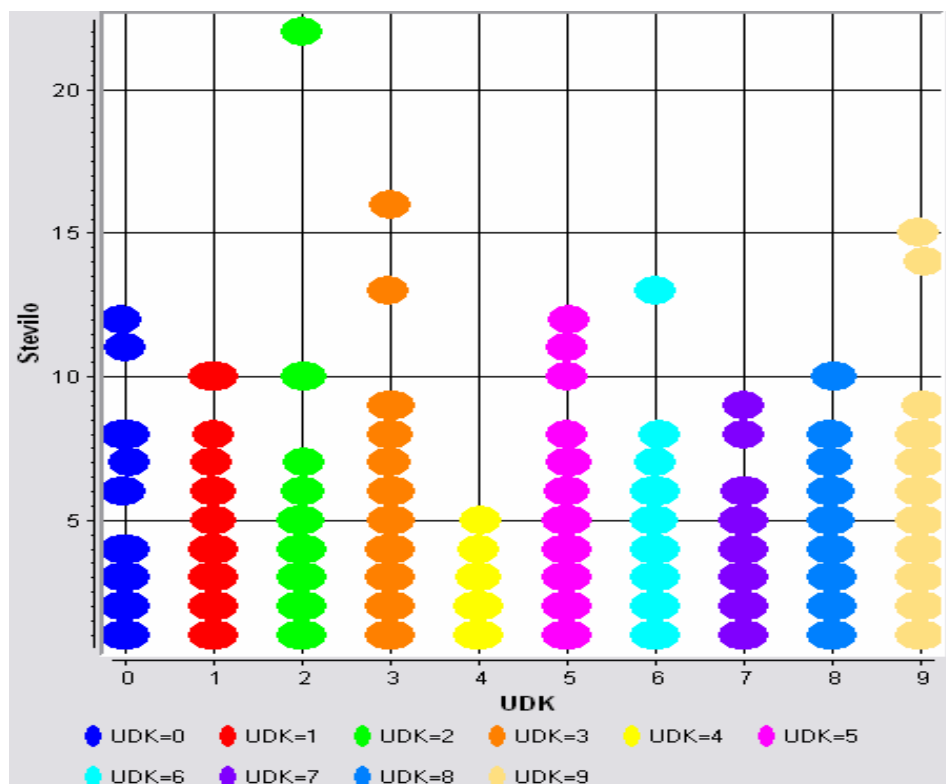
#### 4.5.2 Slikovni prikaz 17: Krivulje verjetnosti in zaupanja znotraj posameznih UDK področij – kakovostni vidik

Slika 17 ponazarja kakovosten pogled na krivulje verjetnosti in zaupanja znotraj posameznih UDK področjih, s pomočjo katerih je možno opaziti določene podobnosti. Prihodnostne napovedi teh krivulj so sorazmerno zvezne, kot da bi imel opraviti z linearnimi trendi, medtem ko kažejo krivulje zanesljivosti ali zaupanja te napovedi

nekoliko manj zvezen izraz. UDK področja, katera so se na podlagi pogostosti poizvedb uvrstila v vrhajo skupino (*UDK 9, UDK 6 in UDK 5*) so si glede obravnavanih krivulj ali intervalov kakovostno gledano zelo podobni. Prav podobno je možno trditi za področje UDK 3, ki spada glede pogostosti poizvedb v srednjo skupino, medtem ko se področji UDK 8 in UDK 7 odmikata od tega vzorca, kajti tako krivulji verjetnosti kot tudi intervali zaupanja prikazujejo neko nezveznost. Obravnavane krivulje ali intervali za področji UDK 1 in UDK 0 (*področji sta uvrščeni v spodnjo skupino*) zopet kažeta na neko zvezno obnašanje, medtem ko pa področje UDK 2 na veliko mero nezveznosti. Krivulje verjetnosti so za naslednja UDK področja padajoče narave: UDK 9, UDK 6, UDK 5, UDK 8, UDK 7 in UDK 1. Krivulja verjetnosti je za področja UDK 0, UDK 2 in UDK 3 naraščajoča. Pri tovrstnem kakovostnem ovrednotenju intervalov nisem vključil še količinske kazalce kot so frekvenčni intervali, dejavniki intervalov verjetnosti in zaupanja na merski skali od 0 do 1.

## 4.6 Scatterplot za število poizvedb po UDK področjih

Izhajajoč iz slike 11 na strani 60 pri poglavju »Priprava podatkov« bom prikazal vizualizacijsko tehniko Scatterplot (*raztreseni diagram*). Primerjal bom med sabo število poizvedb (*Y - os*) po posameznih UDK področjih (*X - os*).



4.6.1 Slikovni prikaz 18: Razpon števila poizvedb znotraj UDK področij

Slika 18 ponazarja razpon števila poizvedb znotraj posameznih UDK področij. Število poizvedb pri področju UDK 0 je enako 809. Spletni obiskovalci so pri določenih poizvedbah poizvedovali po enkrat, dvakrat, trikrat, štirikrat, petkrat, šestkrat, sedemkrat, osemkrat, 11 krat in 12 krat (*gl. temno modri krogi na sliki 18*). Na podlagi Excelove tabele z ustreznimi podatki sem ugotovil, da so spletni obiskovalci v 271 primerih poizvedovali po enkrat, v 137 primerih so poizvedovali po dvakrat, v 46 primerih so poizvedovali po trikrat, v 18 primerih so poizvedovali po štirikrat, v dveh primerih po šestkrat, v enem primeru po sedemkrat, v dveh primerih po osemkrat, v enem primeru po 11 krat in nenazadnje so v enem primeru izvedli poizvedovali po 12 krat (*devet različnih vrednosti*).

Pri področju UDK 1 (*gl. rdeče kroge na sliki 18*) so spletni obiskovalci izvedli 846 poizvedb. Pri tem so v 309 primerih poizvedovali po enkrat, v 152 primerih po dvakrat, v 33 primerih po trikrat, v 17 primerih po štirikrat, v treh primerih po petkrat, v treh primerih po šestkrat, v enem primeru po sedemkrat, v enem primeru po osemkrat in v dveh primerih po 10 krat (*prav tako kot pri UDK 0 devet različnih vrednosti*).

Pri področju UDK 2 (*gl. zelene kroge na sliki 18*) so spletni obiskovalci izvedli 660 poizvedb. Pri tem so v 242 primerih poizvedovali po enkrat, v 116 primerih po dvakrat, v 20 primerih po trikrat, v 10 primerih po štirikrat, v štirih primerih po petkrat, v dveh primerih po šestkrat, v enem primeru po sedemkrat, v treh primerih po 10 krat in v enem primeru po 22 krat (*prav tako kot pri UDK 0 in UDK 1 devet različnih vrednosti*).

Pri področju UDK 3 (*gl. oranžne kroge na sliki 18*) so spletni obiskovalci izvedli 1828 poizvedb. Pri tem so v 586 primerih poizvedovali po enkrat, v 306 primerih po dvakrat, v 88 primerih po trikrat, v 46 primerih po štirikrat, v 15 primerih po petkrat, v štirih primerih po šestkrat, v dveh primerih po sedemkrat, v treh primerih po osemkrat, v treh primerih po devetkrat, v enem primeru po 13 krat in v enem primeru po 16 krat (*11 različnih vrednosti*).

Pri področju UDK 4 (*gl. rumene kroge na sliki 18*) so spletni obiskovalci izvedli 152 poizvedb. Pri tem so v 76 primerih poizvedovali po enkrat, v 26 primerih po dvakrat, v petih primerih po trikrat, v enem primeru po štirikrat in v enem primeru po petkrat (*pet različnih vrednosti*).

Pri področju UDK 5 (*gl. svetlo- vijolične kroge na sliki 18*) so spletni obiskovalci izvedli 1863 poizvedb. Pri tem so v 692 primerih poizvedovali po enkrat, v 330 primerih po dvakrat, v 87 primerih po trikrat, v 26 primerih po štirikrat, v 14 primerih po petkrat, v treh primerih po šestkrat, v dveh primerih po sedemkrat, v dveh primerih po osemkrat, v

enem primeru po 10 krat, v enem primeru po 11 krat in v enem primeru po 12 krat (*11 različnih vrednosti*).

Pri področju UDK 6 (*gl. svetlo- modre kroge na sliki 18*) so spletni obiskovalci izvedli 1917 poizvedb. Pri tem so v 740 primerih poizvedovali po enkrat, v 312 primerih po dvakrat, v 92 primerih po trikrat, v 35 primerih po štirikrat, v devetih primerih po petkrat, v 10 primerih po šestkrat, v enem primeru po sedemkrat in v enem primeru po 13 krat (*devet različnih vrednosti*).

Pri področju UDK 7 (*gl. vijolične kroge na sliki 18*) so spletni obiskovalci izvedli 1155 poizvedb. Pri tem so v 369 primerih poizvedovali po enkrat, v 210 primerih po dvakrat, v 59 primerih po trikrat, v 22 primerih po štirikrat, v 12 primerih po petkrat, v štirih primerih po šestkrat, v enem primeru po osemkrat in v enem primeru po devetkrat (*osem različnih vrednosti*).

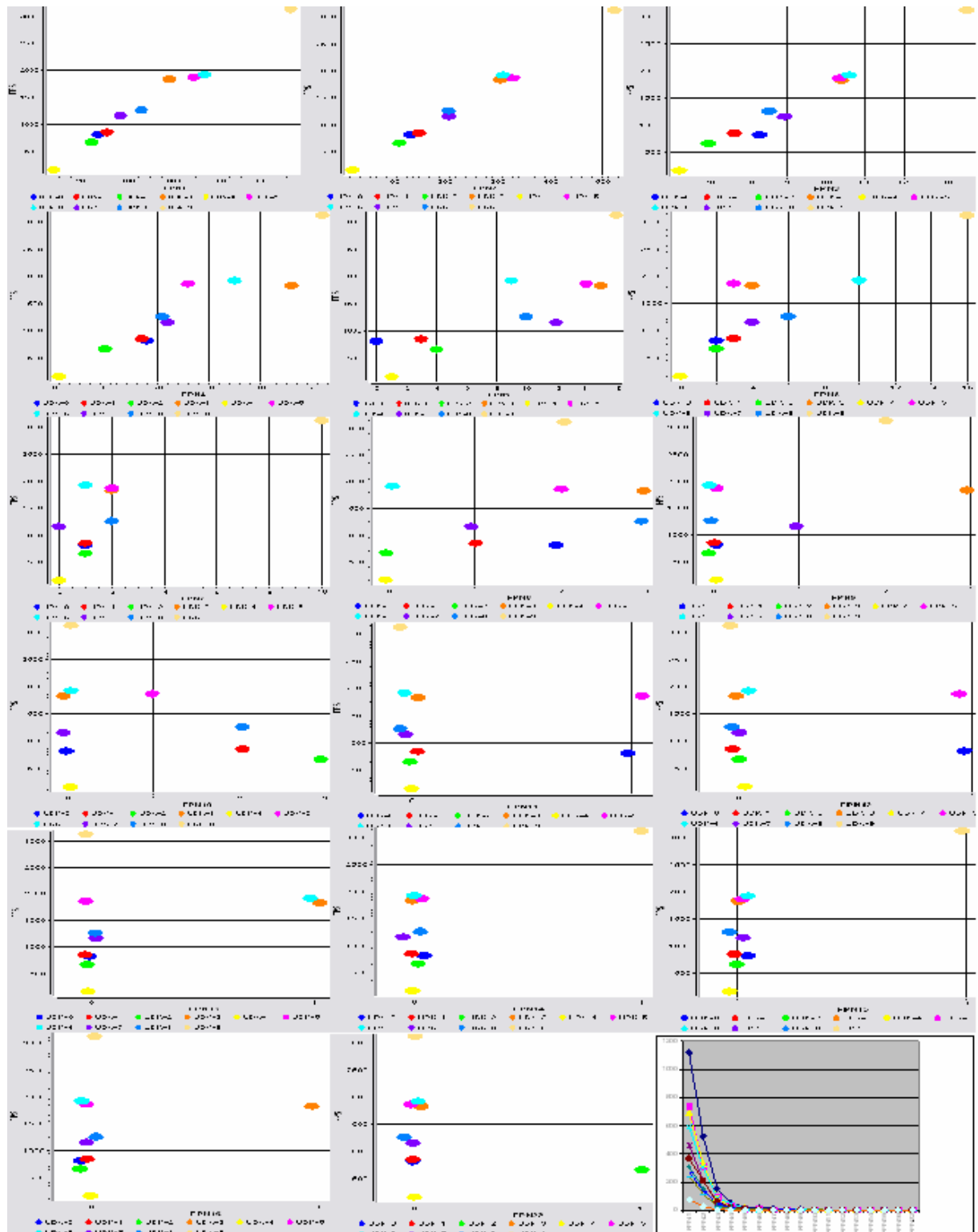
Pri področju UDK 8 (*gl. modre kroge na sliki 18*) so spletni obiskovalci izvedli 1256 poizvedb. Pri tem so v 461 primerih poizvedovali po enkrat, v 208 primerih po dvakrat, v 51 primerih po trikrat, v 21 primerih po štirikrat, v 10 primerih po petkrat, v šestih primerih po šestkrat, v dveh primerih po sedemkrat, v treh primerih po osemkrat in v dveh primerih po 10 krat (*devet različnih vrednosti*).

Pri področju UDK 9 (*gl. svetlo- rumene kroge na sliki 18*) so spletni obiskovalci izvedli 3127 poizvedb. Pri tem so v 1119 primerih poizvedovali po enkrat, v 523 primerih po dvakrat, v 152 primerih po trikrat, v 52 primerih po štirikrat, v 16 primerih po petkrat, v 16 primerih po šestkrat, v 10 primerih po sedemkrat, v dveh primerih po osemkrat, v dveh primerih po devetkrat, v enem primeru po 14 krat in v enem primeru po 15 krat (*11 različnih vrednosti*).

S teh podatkov je možno razbrati, da so spletni obiskovalci najbolj pogosto izvedli eno določeno poizvedbo v 4865 primerih po enkrat, v 2320 primerih po dvakrat, v 633 primerih po trikrat, v 248 primerih po štirikrat, v 84 primerih po petkrat, v 50 primerih po šestkrat, v 20 primerih po sedemkrat, v 14 primerih po osemkrat, v šestih primerih po devetkrat, v osmih primerih po 10 krat, v dveh primerih po 11 in 12 krat in v enem primeru po 13, 14, 15 in 22 krat. Na podlagi tovrstne porazdeljenosti, ki bi jo prikazali slikovno s črtnim diagramom, bi dobili kot izid skoraj pravilno hiperbolično odvisnost (*podobno kot pri frekvenčni distribuciji besed, gl. Zipfov zakon v teoretičnem uvodu*). Prav podobne izide dobimo, če te porazdelitve prikažemo znotraj posameznih UDK področij. Določena poizvedba, ki je bila večkrat zaporedno izvedena in se v poznejšem obdobju več ne pojavlja, v bistvu ni toliko relevantna kot takšna poizvedba, ki se pojavlja po enkrat ali

dvakrat, vendar se pojavlja v različnih časovnih intervalih (npr. od leta 2003 do 2006).

Dokaz za tovrstno trditev je možno najti še zlasti pri poizvedbah s področja UDK 2.



**4.6.2 Slikovni prikaz 19: Porazdelitev razpona poizvedb z ozirom na skupno število poizvedb znotraj posameznih UDK področij in frekvenčna porazdelitev**

Slika 19 prikazuje različne razpone poizvedb (*FPN1 do FPN16 in FPN22*) z ozirom na skupno število poizvedb znotraj posameznih UDK področij (FPS) in frekvenčno porazdelitev obravnavanega razpona (*X – os predstavlja FPN, medtem ko je FPS na Y – osi*). S pomočjo Scatterplota dobimo na vpogled pogostost določene poizvedbe znotraj posameznih UDK področij. Slikovni prikaz 19 je izpeljan iz slikovnega prikaza 18 in ustreznega besednega vpisa, v katerem so bili podani vsi podatki. Ob proučevanju podatkov oziroma slikovnih vzorcev se mi sproži naslednje vprašanje: »Zakaj so nekateri spletni obiskovalci poizvedovali za določen pojem večkrat kot samo enkrat?« Nekateri spletni obiskovalci – poizvedovalci so to storili predvsem iz treh razlogov:

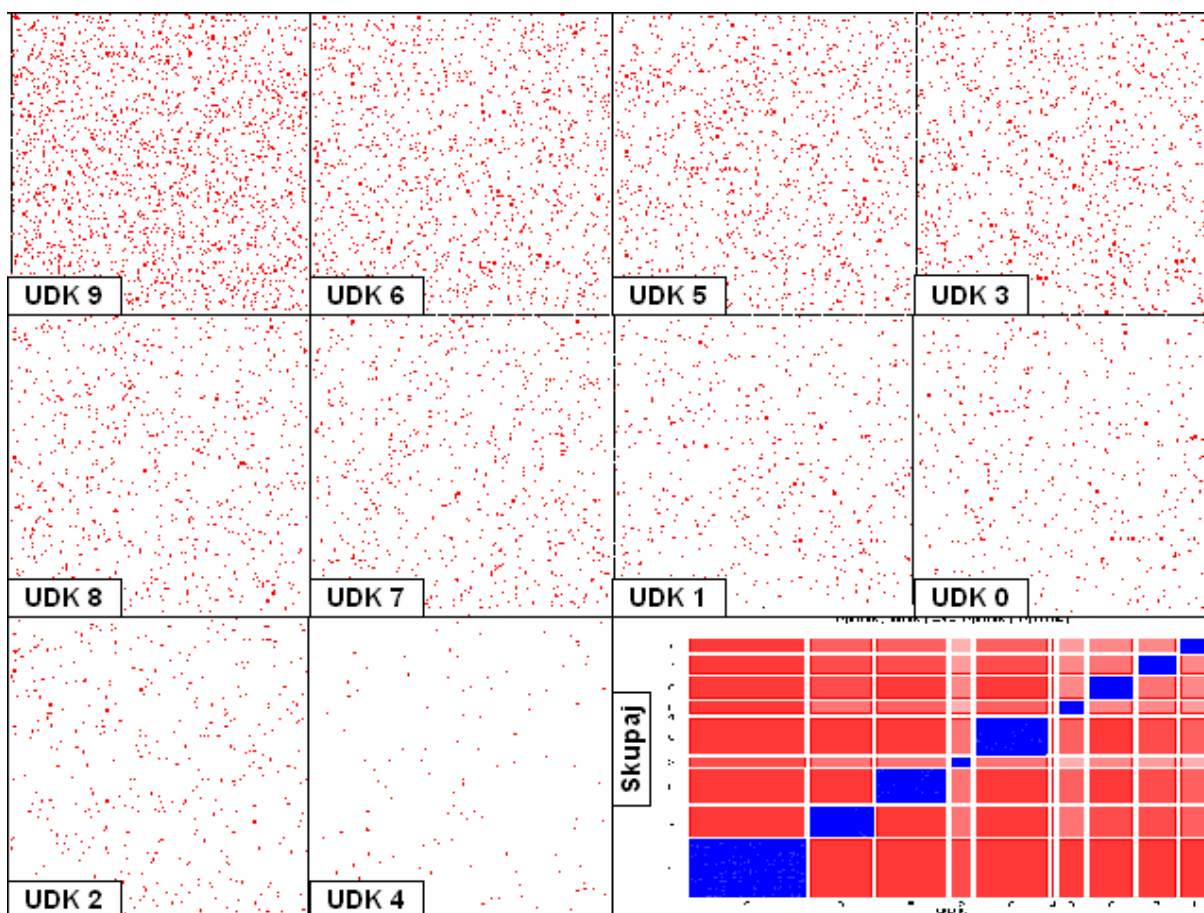
- a.) znotraj izidov so iskali podatke/informacije, ki bi njihovo informacijsko potrebo zadovoljilo, kajti npr. s prvim ali drugim izidom poizvedbe niso bili zadovoljni. Npr. poizvedba »Taliija«, *Evangelijar itd. (izvedena 10 krat)*, *Adam (22 krat izvedena)*, *grški katolik itd.*
- b.) spletni obiskovalci so bili z izidi poizvedbe zadovoljni, vendar so potrebovali dodatne informacije. Npr. *2(100) svetovna verstva, Josip Jurčič Deseti brat (10 krat izvedena)*, *proza, psihologija potrošništva.*
- c.) spletni obiskovalci so z določenim zelo splošnim iskalnim geslom ali ključno besedo preizkušali notranji iskalnik. Poizvedba npr. »Adam« je dajala številne zadetke. Nekaj časa za tem je spletni obiskovalec iskal ime »Eva«, vendar je poizvedoval po tem pojmu zgolj enkrat. Nekateri spletni obiskovalci so tudi uporabljali kratice kot npr. poizvedba »Fr.« (*16 krat je bila ta poizvedba izvedena*).

Od teh treh razlogov ima po mojem mnenju točka c.) najmanjšo težo, sledi mu razlog iz točke b.), medtem ko je razlog iz točke a.) najbolj relevanten! Poizvedbe, ki so se zaporedno zelo pogosto pojavljale, kar pomeni, da je določen spletni obiskovalec to poizvedbo večkrat ponovil ali pa je pregledoval vse izide; so se v poznejšem času izkazale kot relativne »Muhe enodnevnice«. Primeri takšnih poizvedb so že omenjeni »Evangelijar«, Taliija, Adam, grški katolik itd. Zelo pogosta poizvedba kot je npr. »Stres« se je pojavljala v različnih časovnih intervalih povprečno po dvakrat in bi jo lahko imenoval kot »Brezčasno«. Podobno lahko trdim za zelo pogosto poizvedbo »Zgodovina«. Glavna značilnost tovrstnih poizvedb je v tem, da se lahko njihov vsebinski predznak razširja na različne znanosti, umetnosti in športa (*npr. Stres → Stres na delovnem mestu, Stres v transportu, Stres v zdravstvu, Stres v športu ali pa Zgodovina → Zgodovina Slovenije, Zgodovina tenisa, Zgodovina zdravstva, Zgodovina nastanka ekonomskih kriz*). Prikazani slikovni vzorci in frekvenčna porazdelitev (*gl. desno spodaj*

pri sliki 19) poizvedovanja spletnih obiskovalcev kažejo na neko stabilnost in pravilnost, ki je sorazmerno neodvisno od različnih časovnih obdobj.

## 4.7 Prikaz podakovnih vzorcev s pomočjo Sievovega diagrama

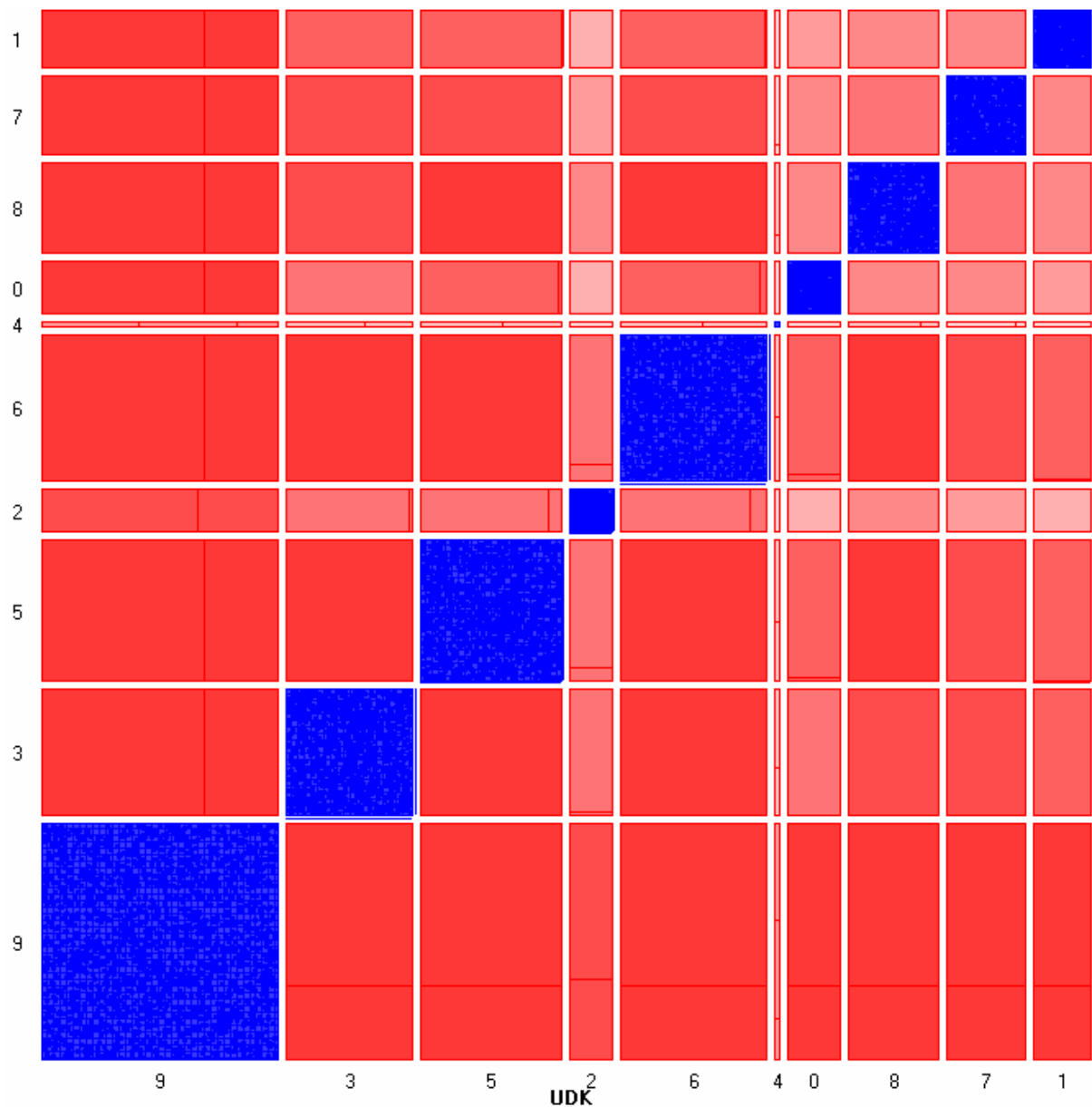
V tem podpoglavju bom uporabil vizualizacijsko tehniko »Sievov diagram« za ponazoritev gostosti poizvedovanja po posameznih UDK področjih (gl. shema na str. 60, slika 11). Temu bo sledil prikaz površine, ki jih zavzemajo posamezna UDK področja (gl. desno spodaj naslednji slikovni prikaz 20).



### 4.7.1 Slikovni prikaz 20: Gostost poizvedb za posamezna UDK področja in skupaj

Slika 20 ponazarja gostost poizvedb po posameznih rangiranih UDK področjih (9, 6, 5, 3, 8, 7, 1, 0, 2 in 10 rang UDK 4) in površino, ki jo v skladu s številnimi poizvedbami zavzemajo področja (gl. desno spodaj sliko 20). Na sliki 20 je možno opaziti največjo gostost poizvedb pri področjih UDK 9, 6, 5 in 3. Pri področjih UDK 8, 7, 1, 0 in 2, pa ta gostost pojema, a pri področju UDK 4 bi prej lahko poročali o relativni redkosti poizvedb. V skladu s temi izidi zavzema področje UDK 9 na skupnem Sievovem diagramu največjo

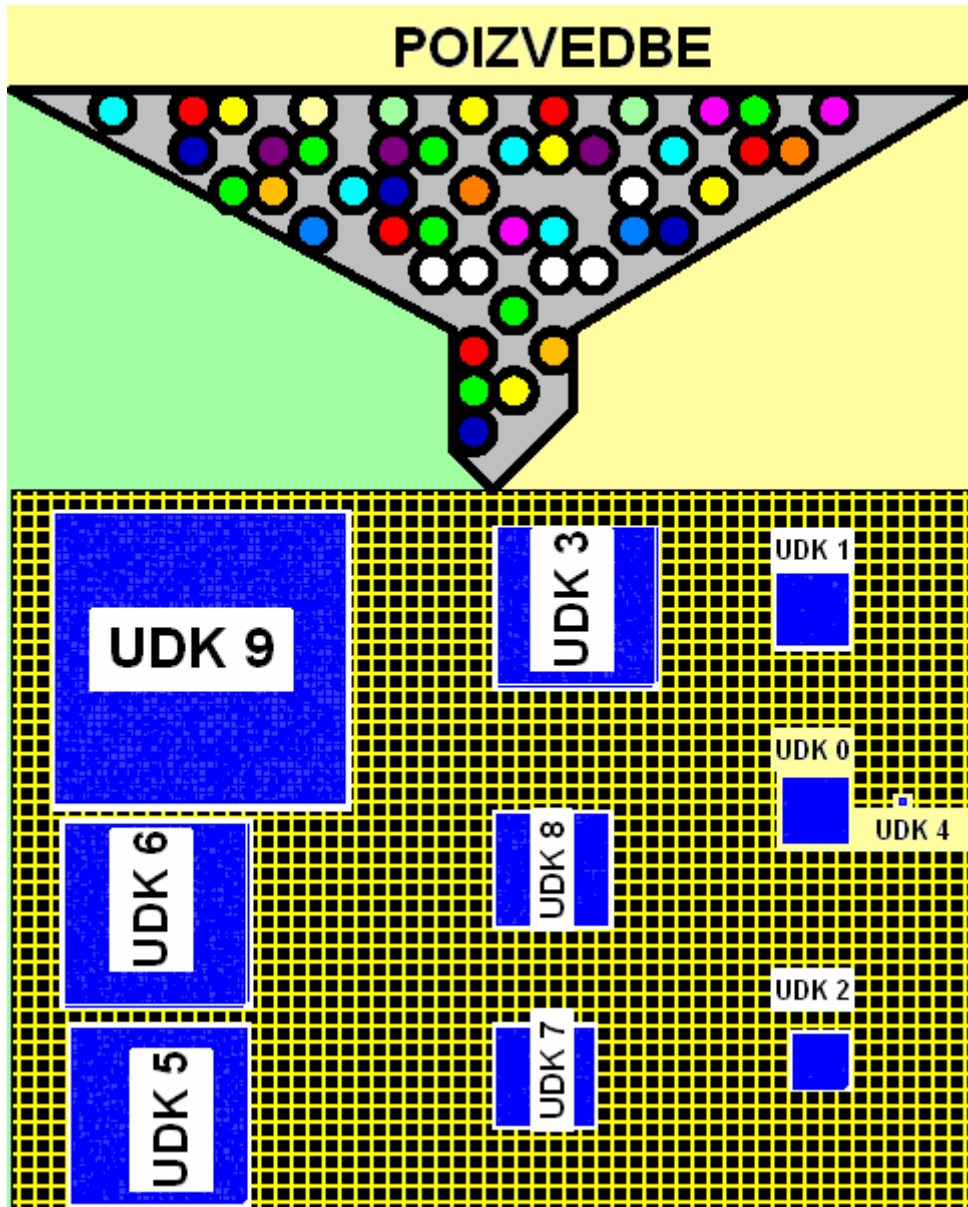
površino. Sledijo mu področje UDK 6, 5, 3, 8, 7, 1, 0, 2 in nenazadnje UDK4. Na ta površinski prikaz se bom na tej strani povrnil in poskušal v nadaljevanju razviti slikovni model zavzete površine po posameznih UDK področjih.



#### 4.7.2 Slikovni prikaz 21: Zavzeta površina poizvedb po posameznih UDK področjih

Slika 21 prikazuje površino, ki jo zavzemajo poizvedbe po posameznih UDK področjih. V tem vpogledu lahko ocenjujem, da zavzemajo največjo površino poizvedbe s področja UDK 9, sledita mu področji UDK 6 in UDK 5, potem UDK 3, UDK 8, UDK 7, UDK 1, UDK 0 in UDK 2. Nenazadnje še lahko poročam o zavzeti površini poizvedb s področja UDK 4 (gl. sliko 21 komajda viden moder kvadrat na poziciji med 6 in 0). Glede na že izveden zakon moči in slikovnega prikaza 21, bi lahko pričakoval, da se bodo v prihodnje

zavzete površine temu ustrezno širile tudi na prostore, ki so do sedaj še ostali nezasedeni (zamisel: teorija o nezasedenem prostoru z ozirom na proizvodbe po posameznih UDK področjih). Zaradi večje nazornosti naj na naslednji strani prikažem obljubljeni slikovni model.



#### 4.7.3 Slikovni prikaz 22: Model sobe oziroma poslopja

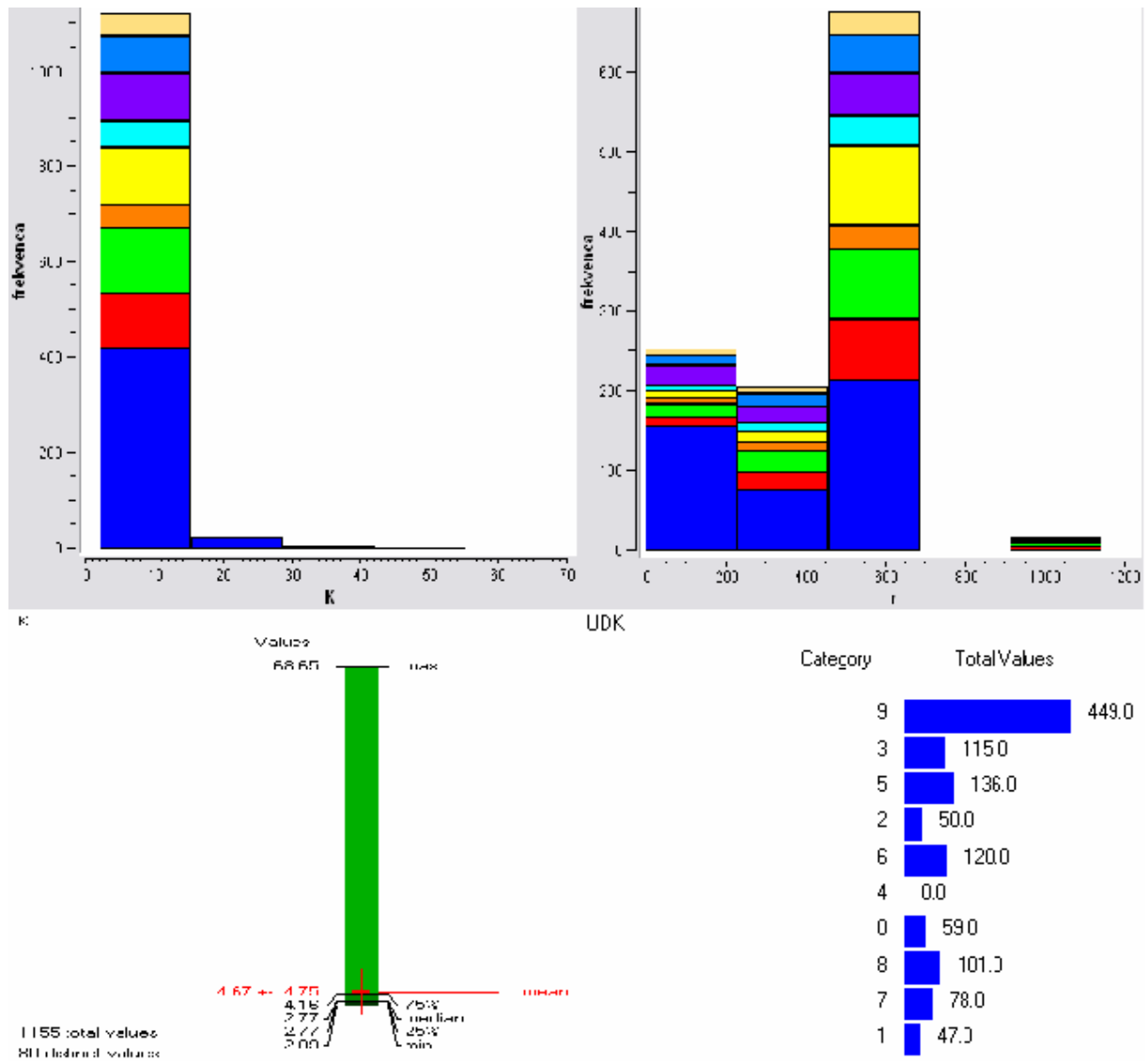
Slika 22 ponazarja površine, ki jih zavzemajo posamezna UDK področja znotraj določenega poslopja. V tem primeru predstavljajo proizvodbe nekakšne vhodne enote, ki se v nadaljevanju razporedijo po posameznih sobah (gl. *slika 22 modri kvadrati*). Ob nadaljnjem pritoku proizvodb, se bodo morale tudi površine posameznih UDK področij temu ustrezno povečati. Črni rumeno obrobljeni kvadratici predstavljajo nezadene prostore določenih proizvodb, ki še niso bile izvedene, vendar pa obstaja sorazmerna velika

verjetnost, da bodo tovrstne poizvedbe v prihodnosti izvedene in razvrščene v ustrezne sobe oziroma UDK področja. S predpostavko, da so sobe sedaj že nasičene ali polne tudi ni več daleč do sklepa, da se bodo morale te sobe razširiti. Sobe se bodo predvsem širile v tisto smer, ki je ustrezna in kjer je dovolj prostora (*podobno kot kristali*). Sobe zaenkrat še imajo pravilno obliko kvadrata, ki pomeni tudi težnjo za prihodnost, vendar pa pravega jamstva, da se bo ta težnja uresničila v bistvu ni, kajti v nekem določenem velikem prostoru so že nastavljeni določeni potenciali, ki sonarekujejo prihodnje dinamike razvoja (*npr. kristal Natrijevega klorida ali kuhinjske soli v generični obliki kocke, se bo razrastle v tisto smer, kjer ima dovolj prostora in tako morda spremenil svojo generično obliko*). Znotraj določene sobe obstajajo prvine, ki so najbolj relevantne, srednje relevantne in manj relevantne. Prav tako so bile nekatere poizvedbe, če izvedem analogijo na predhodni zapis z določenega UDK področja številnejša in vsebinsko pomembnejša od drugih. Znotraj različnih sob obstaja neka hierarhija, ki postavlja temelje za nek določen red in za čim manjšo stopnjo kaosa (*npr. stopnja kaosa je v sobi UDK 4 – gl. majhen komajda viden kvadrater na sliki 22*) največja. Prav tako se nahajajo v sobah prvine, ki so zelo pogosto uporabljene (*ali pa določene poti v sobi so pogosto prehojene, medtem ko druge zelo poredko*). Skratka ta model mi bo v naslednjem poglavju pomagal ovrednotiti dobljene vzorce oziroma izpeljati določena spoznanja in s tem nova znanja o zbranih in obdelanih podatkih.

#### **4.8 Odkrivanje zanimivih vzorcev v podatkih z ozirom na poizvedbe s pomočjo dejavnika pomembnosti ali moči določene ključne besede**

V tretjem poglavju sem prikazal ves postopek za pripravo podatkov in s tem posledično za določitev dejavnika **K** za posamezne pridobljene ključne besede. Zanimive vzorce v podatkih bom prikazal z vizualizacijskimi tehnikami kot so npr. statistični atributi, distribucija in Scatterplotom. Po predstavitvi slikovnega in besednega opisa teh slikovnih tehnik, bom v tem poglavju podrobneje razvil model miselne interesne hierarhije spletnih obiskovalcev/poizvedovalcev. Pri tem se bom naslanjal na že razviti globalni model miselne interesne hierarhije *le-teh* (*gl. sliko 14 na strani 73*). Prav podrobnejši model mi bo v naslednjem poglavju pri izpeljavi novih spoznanj oziroma znanj s podatkov pomagal razviti kasnejši kvalitativni model adaptivne digitalne knjižnice. Naj bo v tem podpoglavju tudi zapisano, da sem sicer izvedel različne in številne preizkuse s pomočjo drugih vizualizacijskih tehnik, ki pa mi v končni stopnji niso dajali ali zadovoljivih in/ali

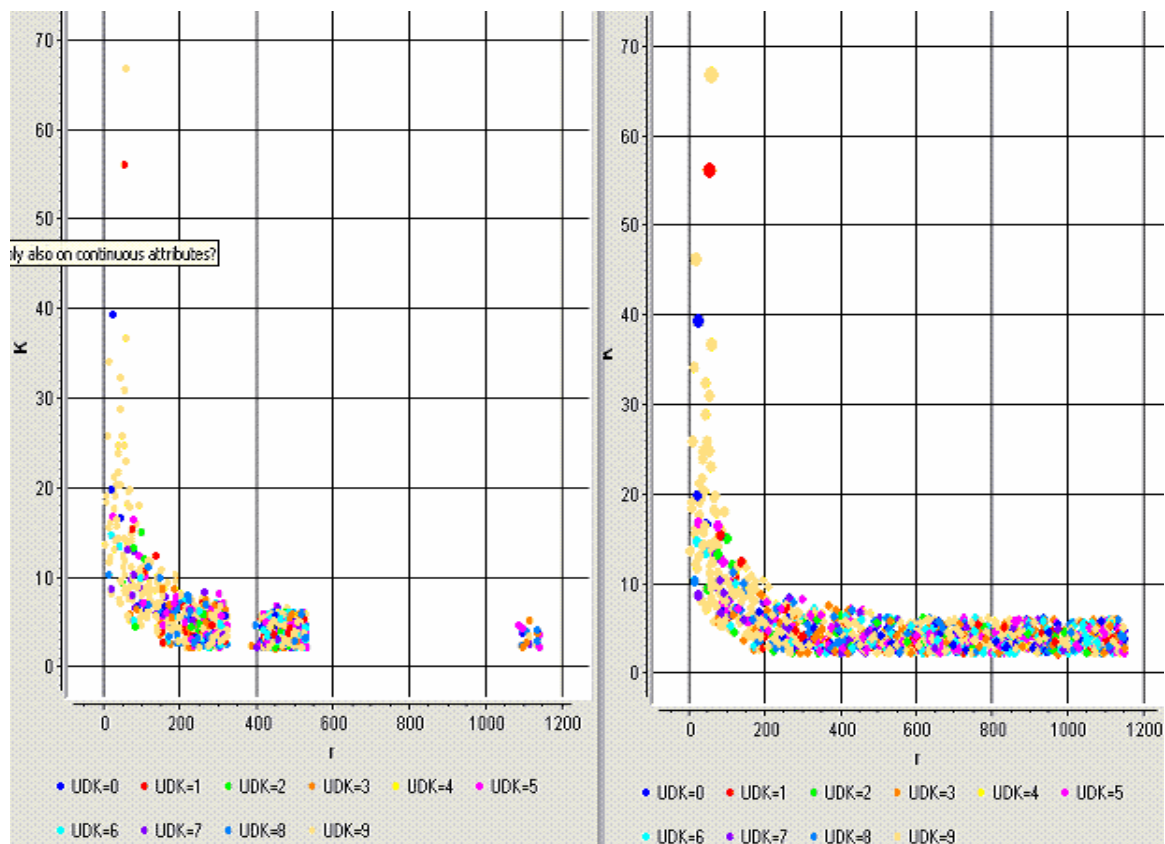
pa razumljivih izidov. Prikazal bom predvsem tiste slikovne prikaze, pri katerih sem primerjal med sabo range in dejavnike **K**. Prav tako nameravam primerjati med sabo določene izide pred in po čiščenju podatkov (*pri tem se bom še zlasti osredotočal na UDK področja*).



#### 4.8.1 Slikovni prikaz 23: Distribucija **K**, ranga, statistični atributi dejavnika **K** in stolpčni diagram UDK

Slika 23 prikazuje distribucijo **K**, ranga, statistične attribute dejavnika **K** in stolpčni diagram za UDK področja. Spodnji del slike predstavlja statistične attribute za dejavnika **K**. Vrednosti **K** se gibljejo od 2,09 do 68,65. Povprečna vrednost je od 4,67 +/- 4,75 in od 1155 vrednosti je 80 različnih. Na spodnjem desnem delu slike je predstavljen stolpčni diagram za UDK področja, kjer pa niso upoštevane vrednosti pogostost pojavljanja določene ključne besede, ki spada v določen UDK razred. Rangirna lestvica se potem ob upoštevanju pogostosti pojavljanja ključne besede, ki spada v neko določeno UDK

področje precej spremeni. Prav to spremembo nam nakazuje upoštevanje ranga ( $r$ ) določene ključne besede. Gornji del slike 23 prikazuje distribucijo dejavnika  $K$  in  $r$ . Sleherno UDK področje je označeno z ustrežno barvo, ki so naslednje: UDK 9 = temno modra, UDK 5 = zelena, UDK 7 = modra, UDK 8 = vijolična, UDK 3 = rdeča, UDK 6 = rumena, UDK 0 = svetlo modra, UDK 1 = svetlo rumena in nenazadnje je področje UDK 2 označeno z oranžno barvo. Področje UDK 4 je bilo s pomočjo dejavnika  $K$  očiščeno. V tem slikovnem prikazu bi želel še zlasti izpostaviti naslednjo zanimivost v zvezi s področjem UDK 7 in UDK 6. Kot je možno s stolpčnega prikaza UDK področij ugotoviti, je področje UDK 6 zasedlo tretje mesto, a področje UDK 7 šesto mesto. Ob upoštevanju dejavnika  $K$  in ranga sta področji zamenjali položaj, kajti ob upoštevanju obeh omenjenih dejavnikov se je izkazalo, da so ključne besede s področja UDK 7 višje rangirane, kar lepo prikazuje zgornji desni del slikovnega prikaza 23. To pomeni, da so bile določene poizvedbe in s tem posledično določene z besedne analize izpeljane ključne besede s področja UDK 7 tako pomenljive in tudi sorazmerno številne, da je prišlo do takšnega preskoka ali zamenjave med področjema UDK 7 in UDK 6. Tako dejavnik  $K$  kot tudi  $r$  bom v naslednjem slikovnem prikazu primerjal med sabo.



**4.8.2 Slikovni prikaz 24: Primerjava dejavnika  $K$  in  $r$  s pomočjo Scatterplota (raztreseni diagram)**

Slika 24 prikazuje primerjavo dejavnika  $\mathbf{K}$  (*Y-os*) in  $\mathbf{r}$  (*X-os*) s pomočjo vizualizacijske tehnike Scatterplota. Pri obeh slikah gre v bistvu za isto snov, le s to razliko, da sem enkrat ključne besede z istimi vrednostmi frekvenc in  $\mathbf{K}$  razvrstil v isti rangirni razred (*gl. levi del slike*), medtem ko sem v drugem primeru sleherno ključno besedo razporedil v poseben rangirni razred (*gl. desni del slike*). S tega je možno spoznati, da dobimo pri levem delu slike nekašne rangirne grozde ali clustre, medtem ko pri desnem delu slike dobimo neko Zipfovo porazdelitev oziroma Paretovo porazdelitev, kjer sorazmerno maloštevilni predstavniki ključnih besed z višjimi vrednostmi dejavnika  $\mathbf{K}$  prispevajo več k skupni vrednosti kot pa sorazmerno množični predstavniki ključnih besed z nižjimi vrednostmi dejavnika  $\mathbf{K}$ .<sup>34</sup> Za poznejšo izpeljavo asociacijskih pravil in grozdov bom uporabil levi del slike 24, medtem ko bom porazdelitev z desne strani slike 24 uporabil za slikovno upodobitev ali vizualizacijo teh podatkov s pomočjo programskega orodja Discovery Ilog (*v tem vpogledu bom tudi slikovno upodobil podatke, ki so vidni v levem delu slike 24*).<sup>35</sup> V prvem primeru imamo opraviti s prekinjeno (*več različnih ključnih besed zaseda isti rang*), medtem ko imamo v drugem primeru opravka z neprekinjeno porazdelitvijo. Glavne značilnosti tovrstne porazdelitve Zipfovega zakona moči so naslednje:

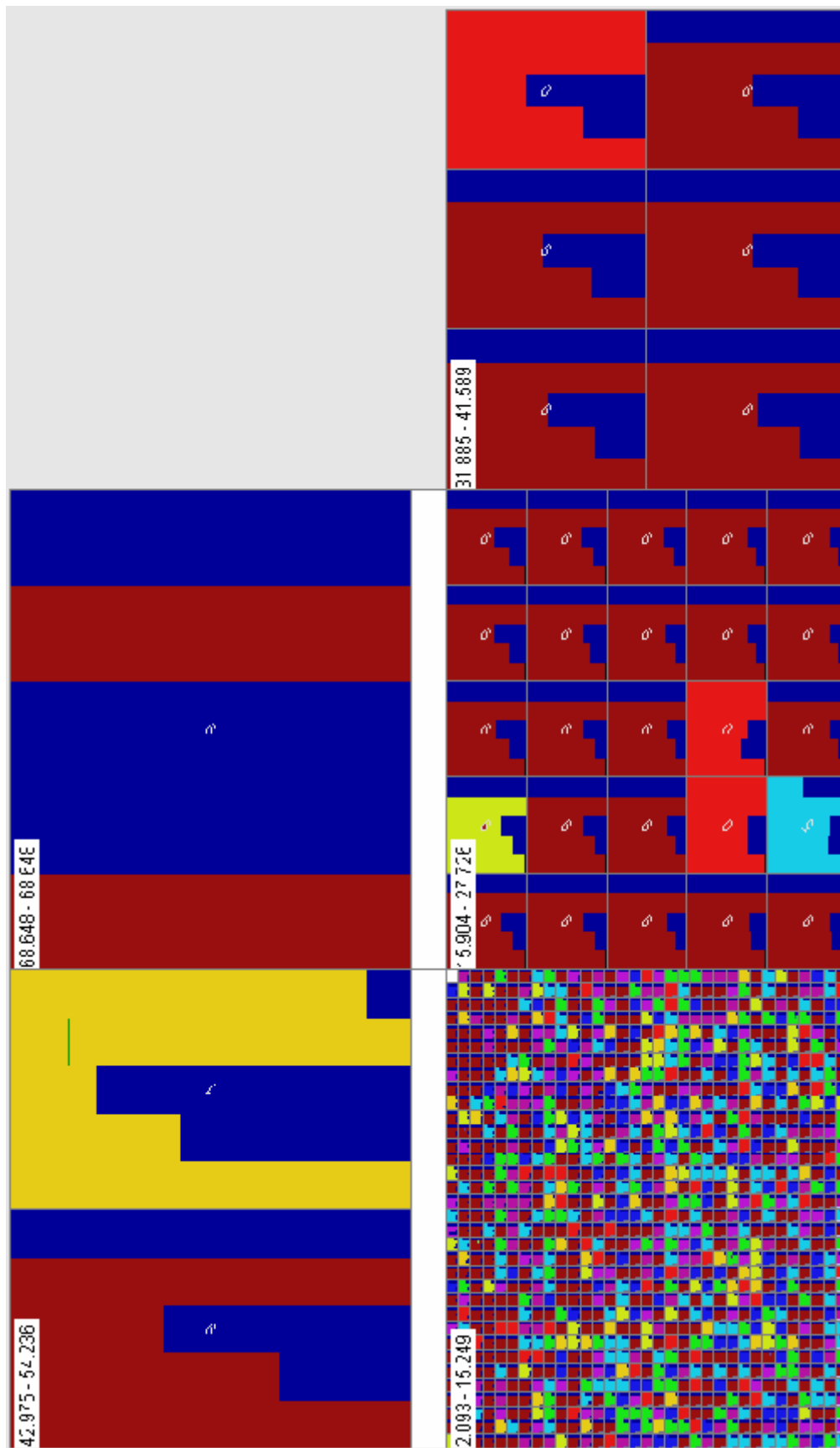
- a. redke prvine dosežejo visoke vrednosti (*vidno na levem predelu diagrama*).
- b. prvine, ki se srednje pogosto pojavljajo, dosežejo navadno srednje visoke vrednosti (*sredina diagrama*).
- c. prvine, ki dosežejo precej nizke vrednosti so navadno manj relevantne (*desna stran Zipfove oziroma Paretove krivulje*).

V primeru te zastavljene raziskave o poizvedbah na notranjem iskalniku UDK leksikona, je možno poročati o besedah, ki so se pojavile precej pogosto (*npr. v, kako, in*), vendar pa njihov pomen ni bil relevanten. Po drugi strani so se pojavljale določene poizvedbe, ki so bile pogoste in so vsebovale močne ključne besede (*npr. zgodovina, stres, Bach*), ki so se pojavljale zelo pogosto ali posamično ali pa v povezavi z drugimi besedami in to z različnih področjih človekovega znanja. Prav tovrstne ključne besede so po izvedeni besedni analizi s pomočjo programskega orodja AntConc-a dosegle visoko vrednost dejavnika  $\mathbf{K}$ . Nekatere druge poizvedbe so bile precej redke, z vidika drugih kontekstov gledano izolirane, tako da so bile te besede po izvedeni besedni analizi ali izločene (*npr.*

<sup>34</sup> Spoznanje izpeljal z naslednjega vira: Barabási, A.L. & Albert, R.(1999). Emergence of Scaling in Random Networks, Science 286, 509.

<sup>35</sup> Programsko orodje je dosegljivo na naslednjem spletnem naslovu: <http://www2.ilog.com/preview/Discovery/index.html> (2006-07-21)

*poizvedbe, ki so se pojavile zgolj enkrat ali pa poizvedbe s področja UDK 4) ali pa so dosegle nizke vrednosti dejavnika K. Prav s tega je možno izpeljati določeno spoznanje, kajti močne ključne besede so lahko interdisciplinarne narave kot npr. »ZGODOVINA« → Zgodovina Evrope, Zgodovina medicine, Zgodovina fizike, Zgodovina glasbe, Zgodovina filozofije, Zgodovina nastanka ekonomskih kriz itd. ali »STRES« → Stres posameznika, Stres na delovnem mestu, Stres in srčna kap, Stres v fiziki itd. Močne ključne besede so v bistvu vsebinsko gledano lahko dvo- ali celo večsmerne kot npr. »BACH« → Življenjepis Johanna Sebastiana Bacha, Delo Bacha, Fuga Bacha, Življenje in delo Bacha itd. V tem primeru je področje UDK 9 izredno močno povezano s področjem UDK 7. Kadar gre za življenjepise znamenitih osebnosti z različnih področij znanosti, umetnosti, športa in različnih dejavnosti, je možno sklepati, da v 95 % primerov obstaja močna povezanost do drugega področja človekovega znanja. Ni naključje, da so spletni obiskovalci mnogokrat poizvedovali po znamenitih osebnostih s področja glasbe, književnosti in svetovne zgodovine - politike. Prav tovrstne poizvedbe so po besedni analizi z AntConc-om dajale ključne besede z visokim dejavnikom **K** kot npr. Bach, Vivaldi, Beethoven, Haydn, Napoleon, Bonaparte, Seliškar, Schubert, Shakespeare, Hitler itd. Poizvedbe v zvezi z boleznijo ali bolezenskimi stanji niso bile redkost (*npr. stres, shizofrenija*). Spletni obiskovalci so tudi večkrat poizvedovali po književnih ali umetniških obdobjih kot npr. klasicizem, romantika, barok itd. Po izvedeni besedni analizi poizvedb so navedene ključne besede dosegle sorazmerno visoke dejavnike **K**. Pri takšnih poizvedbah je možno zapisati, da so večinoma programsko usmerjene, kajti šolarji, dijaki ali študentje morajo za šolo ali fakulteto izdelati domačo nalogo, referat, seminarsko nalogo ali pa diplomsko delo v zvezi s temi temami (*učni program, študijski program*). V bistvu lahko med programsko usmerjene ali programsko spodbujene poizvedbe v 90 % primerih prištevam tudi poizvedbe o znanih književnikih, glasbenikih itd. Dokaj pogoste so bile tudi poizvedbe v zvezi z državami, zato so določene ključne besede kot npr. Slovenija, Grčija, Egipt idr. dosegle sorazmerno visoke vrednosti **K**. Ob tem bi lahko pripomnil, da so bile tovrstne poizvedbe ali programsko spodbujene ali pa individualno spodbujene kot npr. pridobivanje informacij o določeni državi za letovanje ipd., kar je lahko v ožji povezavi s turizmom. Pri poizvedbah, ki sem jih razvrstil v področje UDK 2 kot npr. Yoga, hinduizem bi lahko podobno trdil. Takšne poizvedbe so bile mnogokrat spodbujene tako programsko kot tudi individualno v smeri boljšega psihičnega počutja (*npr. učenje meditacijskih tehnik za boljše počutje*). V naslednjem slikovnem prikazu bom še natančneje odkrival zanimive podatkovne vzorce.*



**4.8.3 Slikovni prikaz 25: Vizualizacija ključnih besed na podlagi dejavnika K**

Slika 25 prikazuje vizualizacijo ključnih besed na podlagi dejavnika **K**. Podatki so razvrščeni po velikosti **K** v pet razredov, znotraj katerih se upoštevajo še drugi kazalci od ranga, frekvence besede in UDK. Barvna skala UDK področij je na sliki naslednja: UDK 0 = rdeča, UDK 1 = rumena, UDK 2 = svetlo rumena, UDK 3 = zelena, UDK 5 = svetlo modra, UDK 6 = temno modra, UDK 7 = svetlo vijolična, UDK 8 = vijolična in nenazadnje UDK 9 je prikazan s temno rdečo barvo. Podatki so prvenstveno razvrščeni in prikazani na podlagi dejavnika **K** v pet različnih velikostnih razredov, ki so naslednji:

1. razred **K** od 68,648 do 68,648: v ta velikostni razred se je uvrstila zgolj ena ključna beseda tj. ZGODOVINA s frekvenco 91 in spada v področje UDK 9.
2. razred **K** od 42,975 do 54,236: v ta velikostni razred se je uvrstila ključna beseda STRES z vrednostjo **K** 54,236 in s frekvenco 53 ter spada v področje UDK 1. Druga ključna beseda, ki se je uvrstila v ta velikostni razred je BACH z vrednostjo **K** 42,975, s frekvenco 31 in spada v področje UDK 9.
3. v razredu s **K** od 31,885 do 41,589 so se uvrstile naslednje ključne besede: UDK z vrednostjo **K** 41,589 (*frekvenca 30, UDK 0*), VIVALDI s **K** 36,044 (*frekvenca 26, UDK 9*), HAYDN s **K** 34,657 (*frekvenca 25, UDK 9*), NAPOLEON s **K** 33,271 (*frekvenca 24, UDK 9*), BEETHOVEN s **K** 33,271 (*frekvenca 24, UDK 9*) in SHAKESPEARE s **K** 31,885 (*frekvenca 23, UDK 9*).
4. v razredu s **K** od 15,904 do 27,726 so se uvrstile ključne besede (*zaradi obširnosti podatkov bom navedel zgolj ključne besede in UDK brez pripadajočih vrednosti*): SELIŠKAR, ANTON, TONE (*vse tri ključne besede spadajo v UDK 9*), SHIZOFRENIJA (*UDK 1*), SCHUBERT, BONAPARTE, GRŠKA, LASSO, JOSEPH, FRANZ, EGIPT, ANTONIO, WILLIAM, SLOVENIJI, ORLANDO, DALI, CHOPIN (*vse UDK 9*), SEMINARSKA, SEMINARSKE (*UDK 0*), IZVOR, RICHARD, PREŠEREN, GRČIJA (*UDK 9*), FOTOSINTEZA (*UDK 5*) in nenazadnje GRŠKI (*UDK 9*).

Kot je bilo možno opaziti v teh velikostnih razredih krepko prevladujejo ključne besede s področja UDK 9. Naslednji velikostni razred bo za besedni opis preobsežen, zato bom zgolj navedel nekatere ključne besede z določenega UDK področja.

5. v razredu s **K** od 2,093 do 15,249 se je uvrstilo največje število ključnih besed z različnih UDK področij. Prav zaradi tega je ta predel slike 25 tudi najbolj barvit. Ta predel slike si je možno s programskim orodjem Ilog Discovery približati oziroma povečati z ukazoma »View« in »Focus In«. V primeru, da si želimo še podrobneje ogledati določen segment enostavno damo še ukaz »Zoom In«. Ko sem ta predel

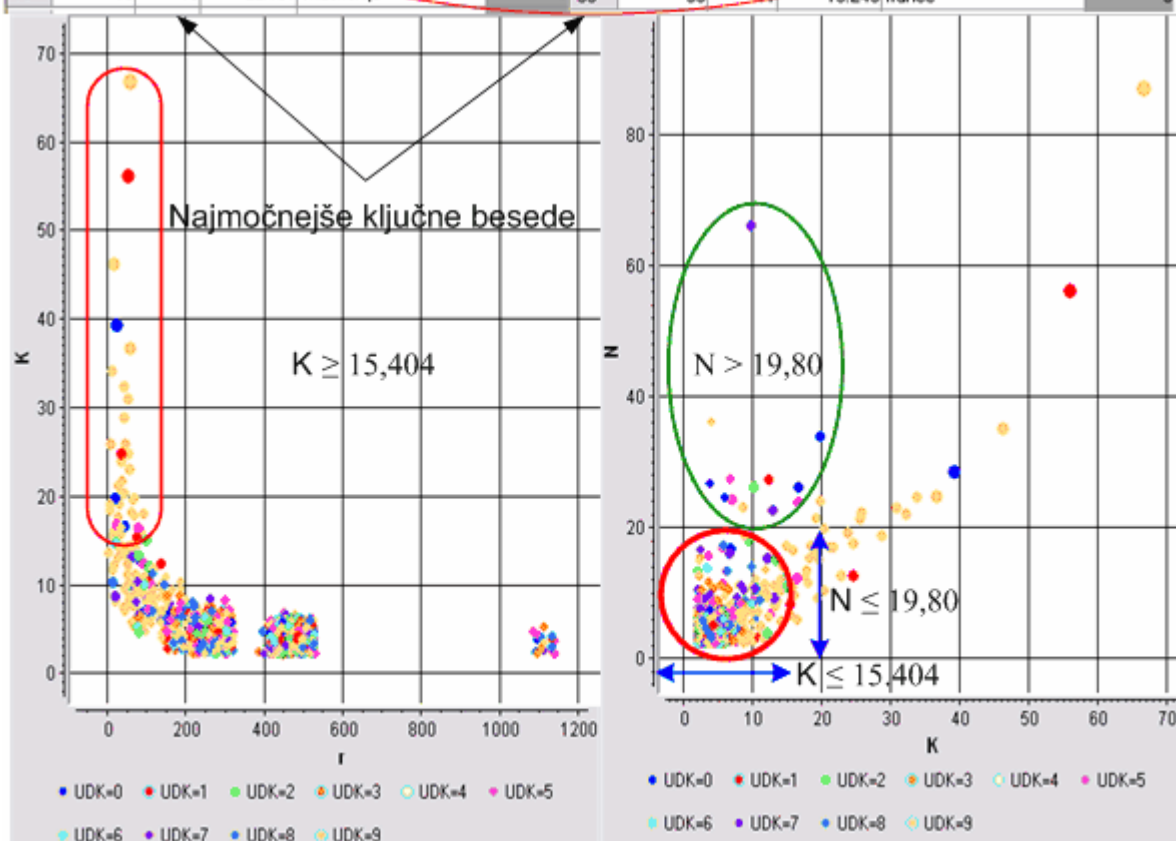
podatkov še dodatno (*to v tem delu ne bom prikazal*) osvetlil se mi je potrdila trditev s prejšnjih strani. Pri globinskem vrtanju v podatke je možno biti še bolj natančen, tako da podrobno preučimo sleherni velikostni razred z ozirom na dejavnika **K**.

Najpomembnejši predstavniki ključnih besed v tem velikostnem razredu so bili naslednji: ZDA (UDK 9), Sebastian (UDK 9), barok (UDK 7), satanizem, hinduizem (UDK 2), klasicizem (UDK 7), ekologija (UDK 5), glasba (UDK 7), Bog (UDK 2), kavčuk (UDK 6), renesanca (UDK 7), Afrika (UDK 9), Islam (UDK 2), angleščina (UDK 8), jehove (UDK 2), Jazz (UDK 7) itd. V tem velikostnem razredu zopet krepko prevladuje področje UDK 9. V tem velikostnem razredu so nekatere ključne besede, ki sem jih razvrstil v področje UDK 7 dosegle sorazmerno visoko vrednost dejavnika **K**, medtem so s področja UDK 6 dosegle nekoliko nižje vrednosti. Z ozirom na pogost pojavljanja UDK področja z upoštevanjem frekvenc sem dobil naslednje izide (*skupaj: 4109*): UDK 9 = 1612, UDK 5 = 480, UDK 7 = 400, UDK 8 = 385, UDK 3 = 362, UDK 6 = 293, UDK 0 = 212, UDK 2 = 202 in nenazadnje s področja UDK 1 je bil izid enak 163. Ta segment je v bistvu posnetek oziroma podoba celostne podatkovne množice izpeljanih ključnih besed, kajti rangirna lestvica je skoraj identična (*področja kot so UDK 1, UDK 0 in UDK 2 so nekoliko zamenjali položaj, če tega primerjamo s celotno podatkovno množico izpeljanih ključnih besed*).

## 4.9 Izpeljana asociacijska pravila

V tretjem poglavju sem že omenil, da gre pri izpeljavi asociacijskih pravil za dva skoraj vzporedna postopka tj. najti ustrezno asociacijsko pravilo z ozirom na pridobljene številčne podatke in ugotavljanje asociacijskih povezav besednih podatkov. Na strani 63 do 64 sem že pripravil asociacijska pravila za številčne podatke glede na dejavnika  $r$  (*ranga*),  $N$  (*frekvence*), **K** in UDK. Po različnih preizkusih in ovrednotenjih sem ugotovil, da je 22. asociacijsko pravilo (*podpora: 0,960; zaupanje: 0,984;  $N \leq 19,80 \rightarrow K \leq 15.404$* ) najbolj ustrezno. To pravilo dobim tudi ob kombinaciji podatkov za  $N$  in **K** ter ob uporabi algoritma za izdelavo induciranih klasifikacijskih pavil (*angl.: Induce Classification Rules*). Ob uporabi tega algoritma dobimo kot izid zgolj eno asociacijsko pravilo. Na podlagi tega asociacijskega pravila sem iz zbirke podatkov z najvišjim dejavnikom **K** izbral 34 ključnih besed (*gl. opis na straneh od 63 do 65*), kajti vrednosti  $N$  in **K** ustrezajo pravilu, nakar sem izdelal .tra datoteko. Naj zaradi večje nazornosti zapisanemu priložim še slikovni in dodatni besedni opis.

r	N	K	B	UDK	16	16	20	21.072	grska	9
1	1	91	68.648	zgodovina	9	17	15	20.794	egipt	9
2	2	53	54.236	stres	1	18	15	20.794	franz	9
3	3	31	42.975	bach	9	19	15	20.794	joseph	9
4	4	30	41.589	udk	0	20	15	20.794	lasso	9
5	5	26	36.044	vivaldi	9	21	14	19.408	antonio	9
6	6	25	34.657	haydn	9	22	13	18.022	chopin	9
7	7	24	33.271	beethoven	9	23	13	18.022	dali	9
8	7	24	33.271	napoleon	9	24	13	18.022	orlando	9
9	9	23	31.885	shakespeare	9	25	13	18.022	sloveniji	9
10	10	20	27.726	anton	9	26	13	18.022	william	9
11	10	20	27.726	seliskar	9	27	30	17.466	seminarska	0
12	12	19	26.340	tone	9	28	23	17.447	seminarske	0
13	13	16	22.181	bonaparte	9	29	20	17.094	izvor	9
14	13	16	22.181	schubert	9	30	12	16.636	grcija	9
15	13	16	22.181	shizofrenja	1	31	12	16.636	preseren	9
					9	32	12	16.636	richard	9
					9	33	22	16.311	fotosinteza	5
					9	34	19	15.904	greki	9
					1	35	11	15.249	france	9



#### 4.9.1 Slikovni prikaz 26: Postopek izpeljave asociacijskih pravil na podlagi dejavnika $K$ , $r$ , $N$ , $UDK$ in ključnih besed

Slika 26 prikazuje postopek izpeljave asociacijskih pravil na podlagi dejavnika  $K$ ,  $r$ ,  $N$ ,  $UDK$  in pomembnih ključnih besed. Levi del slike 26 prikazuje porazdelitev ključnih besed na podlagi primerjave  $r$  in  $K$ , medtem ko desni del slike prikazuje primerjavo med  $K$  in  $N$ . Dobljeno asociacijsko pravilo iz številčnih podatkov po izvedeni besedni analizi

poizvedb (*izhodišče: primerjava med N in K; pogoji: minimalna podpora = 5 %, minimalno zaupanje = 50 %, algoritem = inducirano klasifikacijsko pravilo*) mi je dalo kot izid visoki vrednosti za dejavnika podpore (0,960 ali 96,0 %) in zaupanja (0,984 ali 98,4 %) in povrh tega mi je posredovalo informacijo o tem, na katerem mestu znotraj porazdelitve se nahajajo tiste ključne besede, katerih frekvenca je manjša ali kvečjemu enaka vrednosti 19,80 in katerih **K** so nižji ali kvečjemu enaki vrednosti 15,404 (*gl. rdeče obkrožen barvit podatkovni vzorec v na desni strani slike 26; primerjava med frekvenco in K*). Na podlagi tega asociacijskega pravila za številčne podatke sem lahko izbral najpomembnejše ključne besede, katerih vrednosti **K** so višja ali kvečjemu enaka vrednosti 15,404 (*gl. rdeče podolgovato obkrožen podatkovni vzorec na levi strani slike 26; primerjava med r in K*). Na podlagi tovrstnih najpomembnejših ključnih besed (*gl. gornji del slike 26 v obliki preglednice*) sem potem izdelal obliko .tra datoteko, v kateri so se nahajale te besede. Poleg tovrstnih pomembnih ključnih besed sem v ta dokument uvrstil tudi ključne besede, katerih frekvenca je višja od vrednosti 19,80, a njihove vrednosti **K** se gibljejo od 3,666 do 21,072 (*gl. zeleno obkrožen barvit podatkovni vzorec na desni strani slike 26*). Gre za tiste ključne besede, ki lahko natančneje določijo vsebino najpomembnejših ključnih besed (*gre za ključne besede kot npr. biografija, življenjepis, glasba, bog, leksikon, seminarska, slovar, matematika, psihologija, barok, fotosinteza, grška, delo, služba, transport, književnost ipd.*). Posebnosti pri teh najpomembnejših ključnih besedah so bile npr. naslednje (*nekaj jih bom naštel*): poizvedba »Stres« se je v največjem številu primerov pojavila kot enobesedna zveza in redkeje v dvobesednih ali tro- in več besednih zvezah. Kadar je bila poizvedba v povezavi s stresom dvo- ali več besedna kot npr. »Stres v transportu«, »Stres v zdravstvu« sem tudi tovrstne ključne besede uvrstil v omenjen dokument. Na podoben način sem tudi dopolnil zelo pomembne ključne besede kot so »Bach, Haydn, Vivaldi itd.« (te poizvedbe so *bile v večini primerov enobesedne narave*) npr. s pomožno ključno besedo skladatelj, življenjepis, dela. S programskim orodjem CBA 1.0 za ugotavljanje asociacijskih povezav sem potem izpeljal asociacijska pravila za najmočnejše ključne besede (*postopek je že bil natančneje opisan na strani 62-65*).

V nadaljevanju bom najprej v zgoščeni preglednični obliki prikazal asociacijska pravila za besedne podatke (*izvirno preglednico, ki obsega tri strani bom uvrstil v spremno gradivo tj. na CD-ROM*) in jih slikovno predstavil s programskim orodjem Ilog Discovery, nakar bom dobljene izide po UDK razvrščenih pravil primerjal s pogostostjo UDK področij od mesta 35 do 1155.



sem postopek izpeljave asociacijskih pravil ponovil, pri čemer sem spremenil merila. Kot minimalno vrednost dejavnika **S** sem določil 10 %, medtem ko sem za najmanjšo vrednost dejavnika zaupanja določil 100 %. Kot izid sem dobil sedem asociacijskih pravil, ki so dosegli najvišje vrednosti **S** in **C**:

biografija = Y glasba = Y -> življenjepis = Y (11.429% 100.00% 4 4) UDK 9

biografija = Y -> življenjepis = Y (28.571% 100.00% 10 10) UDK 9

dela = Y življenjepis = Y -> biografija = Y (17.143% 100.00% 6 6) UDK 9

dela = Y biografija = Y -> življenjepis = Y (17.143% 100.00% 6 6) UDK 9

dela = Y -> življenjepis = Y biografija = Y ( 17.143% 100.00% 6 6) UDK 9

dela = Y -> biografija = Y (17.143% 100.00% 6 6) UDK 9

dela = Y -> življenjepis = Y (17.143% 100.00% 6 6) UDK 9

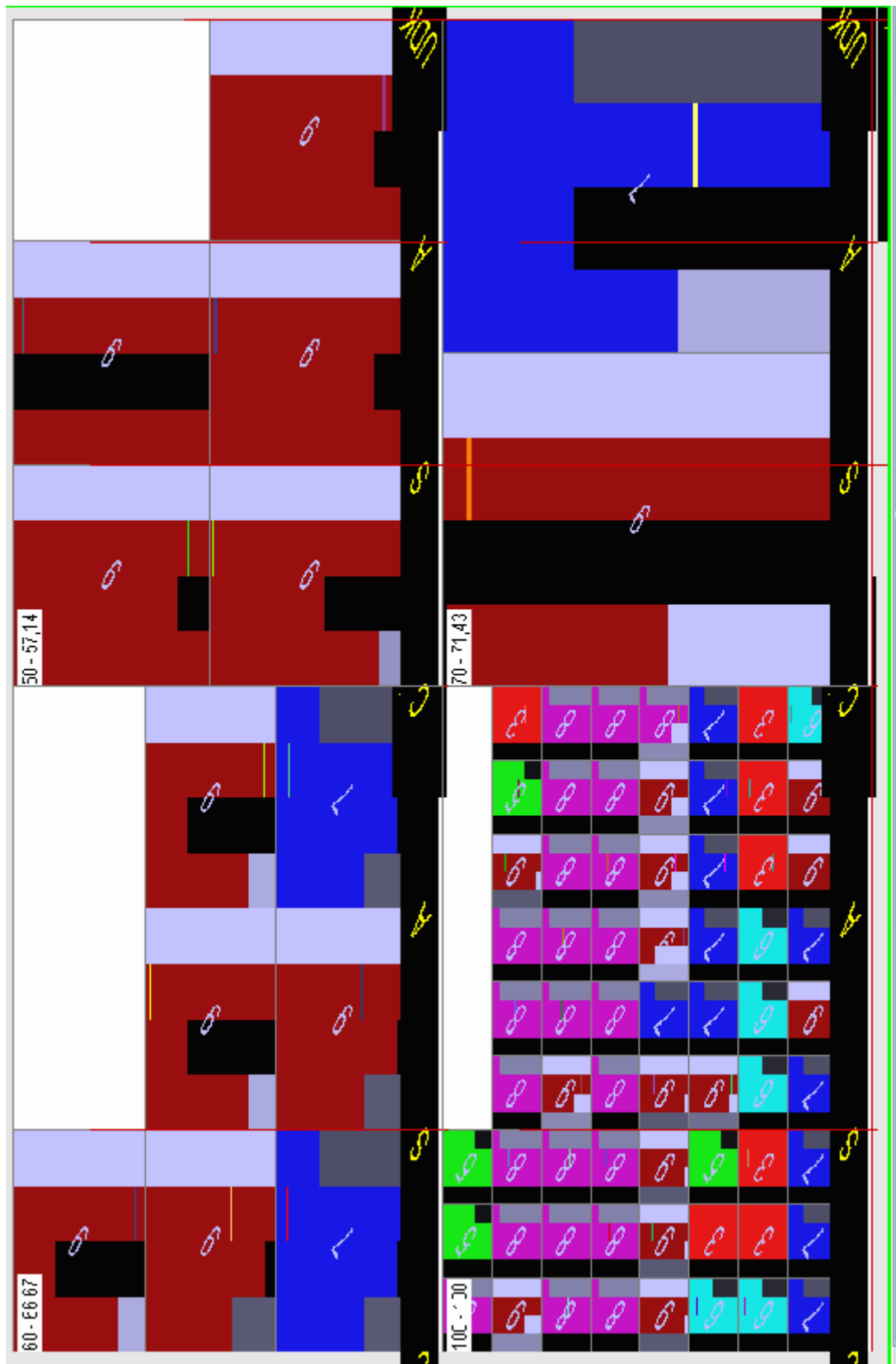
S tega izida lahko sklepam, da so najmočnejša asociacijska pravila s področja UDK 9 ali s področja razvoja človeka in Zemlje, natančneje s področja življenjepisov zgodovinskih osebnosti, ki so bili zelo dejavni in za sabo pustili pomembne sledove ter tako bolj ali manj pomembno vršili vpliv na nadaljnji razvoj človeka in posredno tudi Zemlje. V primeru, da ohranimo merila minimalnega **S** tj. pri 5 % in **C** tj. pri 50 % dobimo zanimiv vpogled v še druge asociacijske povezave, ki povrh tega prikazujejo bolj ali manj močne povezave med različnimi UDK področij, še zlasti so poudarjene povezave med področij UDK 9, UDK 8 in UDK-jem 7.

Dela -> življenjepis, biografija (17,143, 100) UDK 8; Antika, glasba -> zgodovina UDK 7 (5,714, 100); Pesmi, dela, življenjepis -> biografija (5,714, 100) UDK 8; Antika, zgodovina -> glasba (5,714, 100) UDK 7; Pesmi, dela, biografija -> življenjepis (5,714, 100) UDK 8; Glasba, zgodovina -> antika (5,714, 100) UDK 7; Pesmi, življenjepis, biografija -> dela (5,714, 100) UDK 8; Antika -> glasba, zgodovina (5,714, 100) UDK 7 itd. S teh nekaj navedenih pomembnejših asociacijskih povezav s področja UDK 8 in UDK 7 je razvidno, da gre za močno navezo s področjem UDK 9, natančneje s področja življenjepisov znamenitih osebnosti. Poleg tega so opazne tudi močnejše asociacijske povezave med časovnimi obdobji, še zlasti antike (*spada pod zgodovino in ta se nahaja pod UDK 9*) in glasbo (UDK 7) ali književnostjo v smeri pesmi idr. književnih del (UDK 8). Nenazadnje je možno še dodati opažanje, da gre pri teh pomembnejših asociacijskih povezav tudi za predmet zanimanja po skladateljih (*osebnost – življenjepis – UDK 9*) ali književnikih (*osebnost – življenjepis – UDK 9*). Globalni pomen obeh UDK področij je naslednji (*gl. podrobnosti v prejšnjih poglavjih*): UDK 8 pomeni interesno področje

zgradbe in pravil jezika, jezikovne komunikacije, pisanje ter branje o življenju, ki je bilo, je sedaj in bi lahko bilo ipd. V to skupino so se uvrstile poizvedbe s področja jezikoslovja in književnosti. UDK 7 t.j interesno področje kulture, čutnih vtisov in izrazil ter estetike. Če sedaj povežemo globalne pomene področja UDK 9, UDK 8 in UDK 7 lahko dobimo naslednji izrek: »Najmočnejše in najštevilnejše asociacijske povezave s teh treh področij človekovega znanja kažejo na vsebinsko oziroma na področno usmerjenost določenih spletnih obiskovalcev – uporabnikov notranjega iskalnika!« Sledi sinteza: "Razvoj sveta in človeka. Gibanje Zemlje, pojavi, vreme, pomembni zgodovinski dogodki na svetovnih in lokalnih ravneh, znamenite osebnosti in družine, ki so sovplivale na razvoj družbe in s tem morda posredno na sam razvoj Zemlje. Jezik ima pravila in svojo zgradbo in pomeni sredstvo za različne načine sporazumevanja. Sporočanje poteka slušno in pisno. V književnosti se opisuje življenje fiktivno in/ali resnično. Čutna izrazila in estetika soblikujejo življenje oziroma kulturo bivanja. Poteka način komunikacije med izvajalci in poslušalci, ki je lahko slušna, tipna, vizualna ali gibalna. Izredno pomemben je ritem, ki bi ga lahko primerjali z ritmom narave in/ali življenja. **Sklep 1: krajšanje prostega časa, zabava, spoznavanje novih ali manj vsakdanjih stvari. Povpraševanje po izjemnih stvareh, stvaritvah, krajev in ljudi. Sklep 2: v večini primerov so bile poizvedbe s teh področij programsko spodbujene s strani šolskih ali visokošolskih zavodov oziroma programov.**" Ostale asociacijske povezave z drugih UDK področij, ki so sicer manj številne in močne, bom na kratko obravnaval zdaj. Najprej za UDK 6 in UDK 3 (*enotne vrednosti: (5,714, 100)*):

Zdravstvo -> stres; Transport -> stres; Transport, stres -> služba; Služba -> transport, stres; Transport -> služba, stres; Transport -> služba; Služba, transport -> stres; Služba, stres -> transport; Delo -> stres; Služba -> stres; Služba -> transport ; Sociologija -> fizika; Družboslovje -> naravoslovje. Navedene asociacijske povezave se dotikajo problema stresa v družbi in njegovega učinka na zdravstveno stanje ljudi. Najmočnejši negativni kot tudi pozitivni stresni dejavniki so socialnega izvora (*usklajevanje ljudi, razporejanje ljudi, sporazumevanje med ljudmi ipd.*). Socialni stres je v bistvu posledica nekega družbenega programa, ki ne zmore rešiti vseh problemov med ljudmi (*npr. neustrezna komunikacija med ljudmi, neučinkovit transport ljudi na delovna mesta idr.*). **Sklep: programsko reševanje različnih prvin v družbi (zdravje ljudi, oskrbovanost ljudi z različnimi dobrinami, promet idr.).** Za področje UDK 5: Fizika -> sociologija; Naravoslovje -> družboslovje; Rastline -> biologija; Biologija -> rastline (*vrednosti: 5,714, 100*). **Sklep: odnos med družbo in naravo, med ljudmi in živimi bitji.**

S pomočjo naslednjega slikovnega prikaza bom poskusil še druge podrobnosti izluščiti.

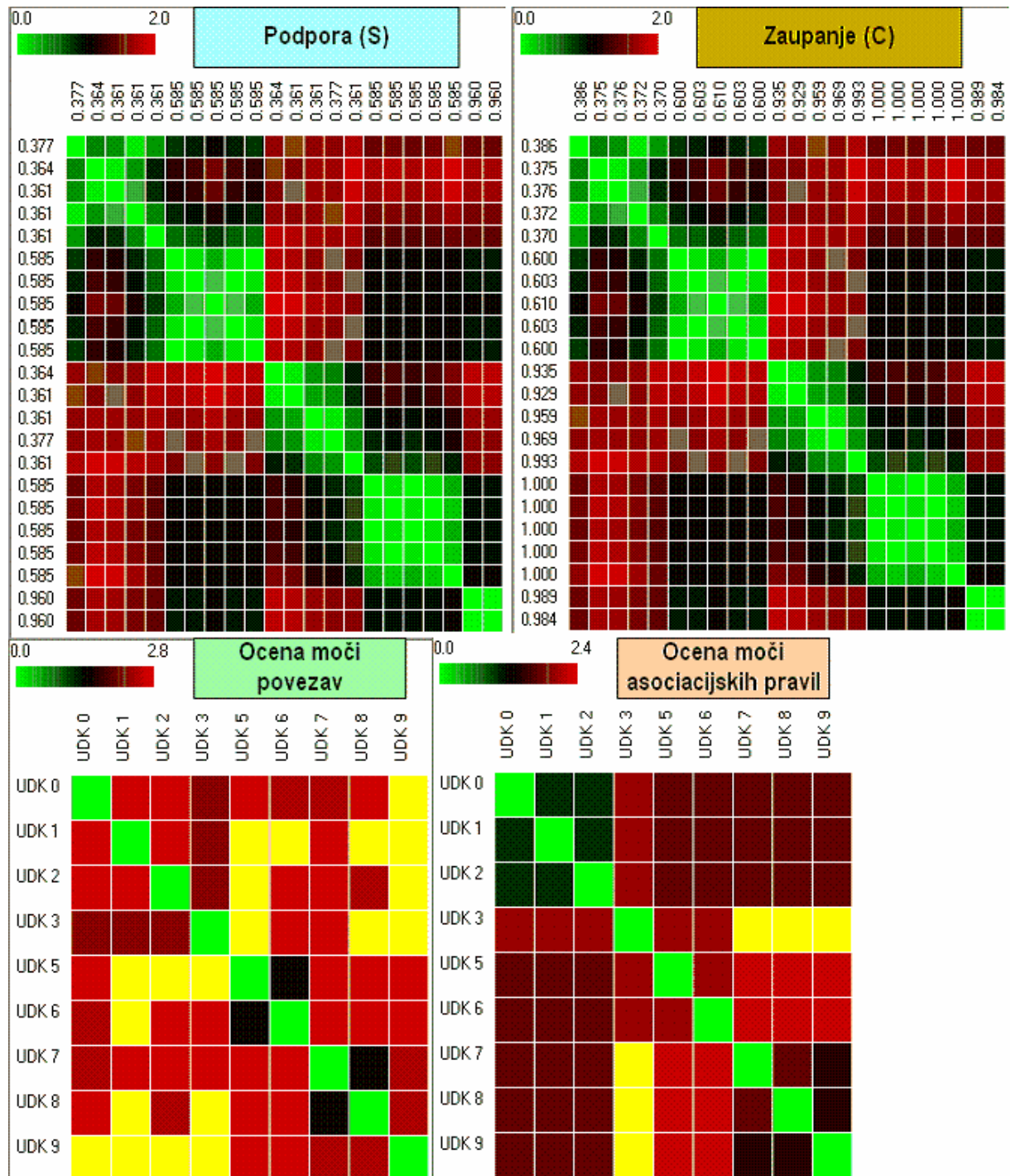


**4.9.3 Slikovni prikaz 27: Vzorec porazdelitve asociacijskih povezav**

Slika 27 prikazuje vzorec porazdelitve asociacijskih povezav na podlagi dejavnika C (gl. sliko 27: razdeljenost na štiri frekvenčne razrede po C tj. 100 %, od 70 % do 71,43 %, od 60 % do 66,67 % in od 50 % do 57,14 %) in znotraj tega po S, A in še zlasti po UDK področjih (gl. številčne zaznamke na sliki 27). Nemudoma je možno opaziti, da je razred C (100 %) najbolj raznolik vzorec. V razredih z nižjim dejavnikom C (od 50 % do 100 %) se asociacijske povezave s področja UDK 9 pojavijo 11 krat, medtem ko se asociacijske povezave s področja UDK 7 pojavljajo trikrat. V razredu z najvišjo vrednostjo C (100 %) je sestava glede na pogostost pojavljanja asociacijskih povezav po UDK področjih naslednja: UDK 9 = 14, UDK 8 = 24, UDK 7 = 11, UDK 6 = 6, UDK 5 = 4 in UDK 3 = 7. Skupni izid je naslednji (podal bom odstotkovne deleže gleda na skupno število asociacijskih pravil, ki je 80):

UDK 9 =  $25 \times 100/80 = 31,25 \%$ ; UDK 8 =  $24 \times 100/80 = 30,00 \%$ ; UDK 7 =  $14 \times 100/80 = 17,50 \%$ ; UDK 6 =  $6 \times 100/80 = 7,50 \%$ ; UDK 5 =  $4 \times 100/80 = 5,00 \%$  in UDK 3 =  $7 \times 100/80 = 8,75 \%$ . Odstotkovna sestava ključnih besed od mesta 35 do 1155 po posameznih UDK področjih je naslednja: UDK 9 =  $1612 \times 100/4021 = 40,09 \%$ ; UDK 8 =  $374 \times 100/4021 = 9,30 \%$ ; UDK 7 =  $389 \times 100/4021 = 9,67 \%$ ; UDK 6 =  $282 \times 100/4021 = 7,01 \%$ ; UDK 5 =  $469 \times 100/4021 = 11,66 \%$ ; UDK 3 =  $351 \times 100/4021 = 8,73 \%$ ; UDK 2 =  $191 \times 100/4021 = 4,75 \%$ ; UDK 1 =  $152 \times 100/4021 = 3,78 \%$ ; UDK 0 =  $201 \times 100/4021 = 5,00 \%$ . Asociacijskih pravil ni bilo s področja UDK 0, UDK 1 in UDK 2. Še zlasti nekoliko preseneča ugotovitev, da ni bilo s področja UDK 1 nobenih asociacijskih pravil, čeprav je bilo možno na prvih 34 mestih beležiti dve močni ključni besedi kot sta »Stres« in »Shizofrenija«, ki sta se pogostokrat kot poizvedbi pojavili kot enobesedna zveza, kar spada pod UDK 1, vendar v vsebinski relaciji sta se pojavili v povezavi UDK 3 in še zlasti s področjem UDK 6. Ključna beseda »Stres« se je pojavila še zlasti v zvezi s službo, transportom, delovnim mestom in zdravstvom, medtem ko se je ključna beseda »Shizofrenija« pojavila zgolj kot enobesedna zveza in tako zgolj posredno odpirala povezavo na asociacijo psihologija, bolezen, in zdravstvo. Odstotkovni delež ključnih besed za področja UDK 0, UDK 1 in UDK 2 od mesta 35 do 1155 je nižji in se giblje od 3,78 % do 5,00 %, medtem ko je odstotkovna sestava za asociacijska pravila za ta področja enaka 0 %. Področja kot so UDK 9, UDK 8, UDK 7, UDK 6, UDK 5 in UDK 3 so glede zastopanosti ključnih besed od mesta 35 do 1155 dosegla z ozirom na odstotkovne deleže višje izide tj. od 7,01 % do 40,09 %. Prav v vpogledu z asociacijskimi pravili je možno opaziti neko povezavo, kajti izpeljana asociacijska pravila so bila razvrščena pod UDK 9, UDK 8, UDK 7, UDK 6, UDK 5 in UDK 3.

Na podlagi doslej dobljenih izidov bom lahko v nadaljnji stopnji izdelal distančne mape za dejavnika C in S v zvezi s številčnimi podatki, prikaz moči povezanosti med UDK področij z ozirom na poizvedbe in ključne besede ter dejanski prikaz moči dobljenih asociacijskih pravil, ki so bila razvrščena po UDK.



**4.9.4 Slikovni prikaz 28: Distančne mape za dejavnika podpore in zaupanja ter ocenjena moč asociacijskih povezav in pravil**

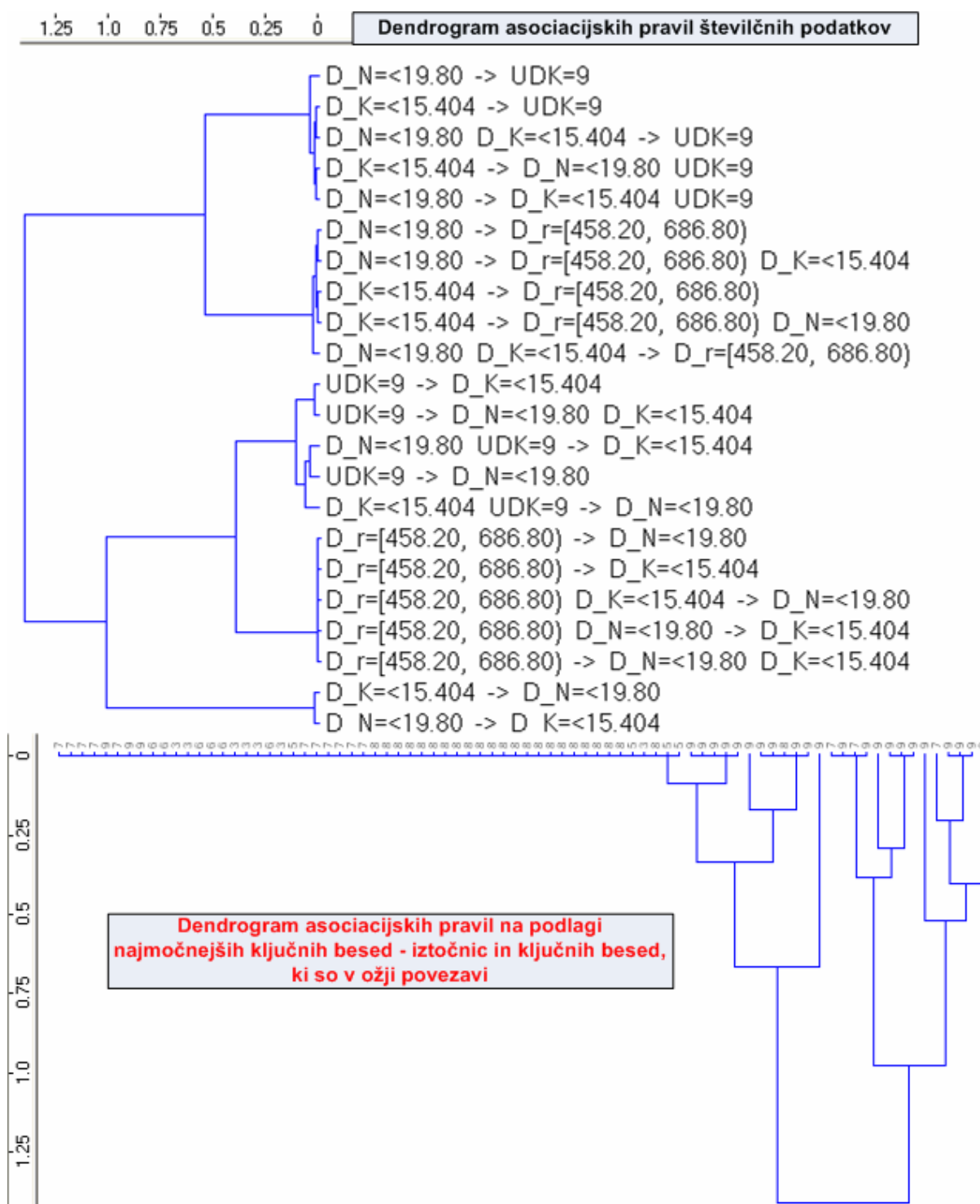
Slika 28 prikazuje dve distančni mapi za dejavnika podpore in zaupanja (*merilo: evklidska razdalja – gl. gornji del slike 28*) ter dve distančni mapi za prikaz ocenjene moči asociacijskih povezav in pravil (*merilo evklidske razdalje – gl. spodnji del slike 28*).

Gornji del slike 28 prikazuje barvne vzorce evklidskih razdalj med dvema vrednostima tako za dejavnika podpore kot tudi zaupanja na lestvici od 0,0 do 1,0 (*najmanjša evklidska razdalja – prikazana s svetlo zeleno, zeleno in temno zeleno barvo*), od 1,0 do 1,067 (*prikazano s črno obarvanimi kvadrati*) in nenazadnje od 1,067 do 2,0 (*največja evklidska razdalja, kar je prikazano s prehodom od temno rjavih, rjavih pa tja do rdečih kvadratov*). Najmanjše evklidsko razdalje med dvema vrednostima je možno opaziti na desnem spodnjem delu gornje slike 28, ki znaša (*ob S1 = 96,0 %, C1 = 98,9 %; S2 = 96,0 %, C2 = 98,4 %*) 0,0121850725. Gre za že znani izpeljani asociacijski pravili izpeljani iz številčnih podatkov po izvedeni besedni analizi poizvedb  $D_K \leq 15.404 \rightarrow D_N \leq 19.80$  (96,0 %, 98,9 %) in  $D_N \leq 19.80 \rightarrow D_K \leq 15.404$  (96,0 %, 98,4). V območju  $K \leq 15,404$  in  $N \leq 19,80$  se nahaja največje število ključnih besed (*gl. sliko 26*), ki pa kot sem že ugotovil glede dejavnika K, r in N niso najmočnejša. Prav s to ugotovitvijo sem lahko pri izpeljavi najizrazitejših asociacijskih pravil z ozirom na najpomembnejše ključne besede izločil tiste ključne besede, ki so se glede vrednosti K in N nahajali v razponu  $K \leq 15,404$  in  $N \leq 19,80$ . Ob tej uporabljeni metodi, ki je bila glede asociacijskih pravil izločilne narave, bo možno v nadaljevanju lažje določiti hierarhične grozde različnih skupin ključnih besed, ki so razvrščene v različna UDK področja. Preden bom lahko to izpeljal je potrebno še najprej obrazložiti spodnji del slikovnega prikaza 28. V levem spodnjem delu slike 28 je prikazan vzorec moči asociacijskih povezav, kjer sem predhodno ovrednotil moč asociacijskih povezav med posameznimi UDK področij na velikostni skali od 0, 1, 2, 3, 4, 5 (*0 – najšibkejša stopnja povezanosti, 5 – najmočnejša stopnja povezanosti*) z vidika izida analize poizvedb in iz njih izpeljanih pomembnih ključnih besed. Evklidske razdalje ovrednotenih asociacijskih povezav med posameznimi UDK področij so označena s pomočjo barvne skale od 0,0 do 2,8 (*zelena = najmanjša evklidska razdalja, črna = manjša evklidska razdalja, rjava = večja evklidska razdalja, rdeča = največja evklidska razdalja, medtem ko pomenijo rumeno označena polja prekoračitev vrednosti tj. nad 2,8*). Na podlagi dobljenega barvnega vzorca oziroma vrednosti evklidskih razdalj je možno trditi, da so najmanjše evklidske razdalje med UDK 6 in UDK 5 ter UDK 8 in UDK 7 tj. 1,414. Na drugem mestu glede najmanjših evklidskih razdalj so se uvrstile ovrednotene asociacijske povezave med UDK 9 in UDK 8 ter med UDK 9 in UDK 7 (*evklidska razdalja je enaka 2,4494...*). Desni spodnji del slike

28 prikazuje barvit vzorec evklidskih razdalj na lestvici od 0,0 do 2,4. V tem primeru sem na isti velikostni skali ocenil moč povezave po UDK razvrščenih asociacijskih pravil. V tem vpogledu lahko zanemarimo vsa polja, ki so ovrednotena z vrednostjo 0, še zlasti od UDK 0, UDK 1 in UDK 2, kajti s teh področij kot izid nisem dobil niti eno asociacijsko pravilo. Prav tako je možno zanemariti z rumeno barvo označena polja. S tega izhajajoč dobimo najmanjšo evklidsko razdaljo za asociacijska pravila med UDK 9 in UDK 8, med UDK 9 in UDK 7 (*vrednost je 1,414...*). Vrednost evklidske razdalje je enaka 2,0 dobimo za asociacijska pravila med UDK 6 in UDK 5. Moč povezave med asociacijskimi pravili s področja UDK 8 in UDK 7 je enaka 1,732. Nenazadnje obstaja še ena ovrednotena moč relacije za asociacijska pravila med UDK 6 in UDK 3 (*vrednost je enaka 2,0*). Glavno središče moči povezanosti med asociacijskimi pravili po razvrščenih UDK področij se nahaja pri UDK 7, 8 in 9, medtem ko se manjše središče nahaja pri UDK 3, UDK 5 in UDK 6. Prav ta ugotovitev je pomembna, ker mi daje vedeti, da se na teh področjih nahajajo tudi največji hierarhični grozdi. Prav s hierarhičnim grozdenjem se bom ukvarjal v naslednjem podpoglavju.

#### **4.9.5 Hierarhično grozdenje asociacijskih pravil in ključnih besed**

Pričujoče podpoglavje bo prikazovalo hierarhične grozde asociacijskih pravil številčnih in besednih podatkov ter ključnih besed s pomočjo slikovne tehnike dendrograma, ki je bil predstavljen že v teoretičnem uvodu. Hierarhične grozde bom izdelal na podlagi postopka, ki je bil slikovno in besedno predstavljen na strani od 63 do 65 (*gl. slikovni prikaz 12 na strani 63 in nadaljnji besedni opis: merilo bo evklidska razdalja za posamezna UDK področja, kot merilo za vrsto prepletenosti ali povezanosti bom določil povprečno – average linkage*). V določeni meri je bilo že s pomočjo ugotavljanja najmočnejših asociacijskih pravil možno ugotoviti oziroma deloma razčleniti glavna zanimanja spletnih obiskovalcev po posameznih področjih človekovega znanja. Razčlemba zanimanja spletnih obiskovalcev pa v bistvu posredno ali bolje deloma tudi razčlenjuje spletne obiskovalce oziroma poizvedovalce v posamezne skupine. Izpeljana asociacijska pravila so mi dajali informacijo o glavnih podatkovnih grozdih, zanemarjali so pa tiste podatkovne množice, ki so lahko za razčlemba poizvedovalcev tudi zelo pomembne (*tovrstne podatkovne gruče bi lahko slikovito primerjal s sekundarnim rečnim tokom, ki sicer ni prisoten v sredini velike reke, ampak kljub temu pomeni nekakšna podstava za glavni rečni tok in s tem seveda sooblikuje obliko in gibanje reke oziroma rečnega toka*).



#### 4.9.6 Slikovni prikaz 29: Dendrograma hierarhičnih grozdov za asociacijska pravila

Slika 29 ponazarja dendrograma hierarhičnih grozdov za asociacijska pravila. Gornji del slike 29 prikazuje hierarhične grozde asociacijskih pravil, ki sem jih izpeljal na podlagi številčnih podatkov najpomembnejših ključnih besed po besedni analizi poizvedb. Opaziti je možno, da je dendrogram najbolj razvejen pri zadnjih dveh asociacijskih pravil ( $D_K \leq 15,404 \rightarrow D_N \leq 19,80$  in  $D_N \leq 19,80 \rightarrow D_K \leq 15,404$ ), kajti veje vozlišča se

povezujejo z drugimi vejami. S tega opažanja je možno izpeljati sklep, da predstavlja to asociacijsko pravilo na podlagi primerjave med N in K tudi poglobitni grozd. Grozd vsebuje 12 prvin in višine grozdov po hierarhičnih nivojih so naslednja:

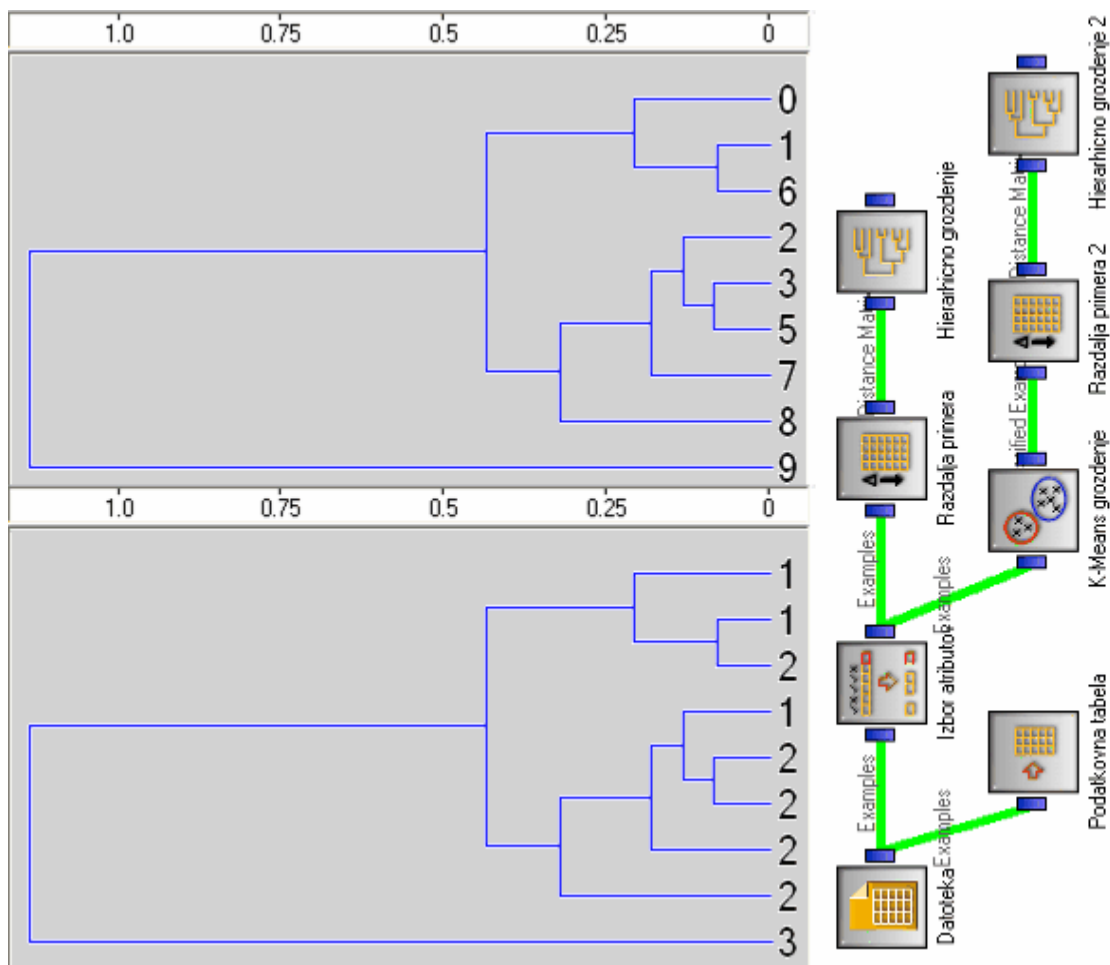
1. hierarhični nivo grozda: višina je enaka 0,007937.
2. hierarhični nivo grozda: višina je enaka 1,004525
3. hierarhični nivo grozda: višina je enaka 1,401922

Skratka zadnji dve asociacijski pravili najbolj označujeta poglobitni hierarhični grozd. Spodnji del slike 29 prikazuje dendrogram asociacijskih pravil, ki sem jih izpeljal na podlagi izbora najmočnejših ključnih besed – iztočnic in ključnih besed, ki so z njimi najbolj povezani. Najbolj razvejeno zgradbo dendrograma je možno zaslediti na desni strani slike 29. Gre za asociacijska pravila, ki sem jih razvrstil še zlasti v področje UDK 9, UDK 8 in UDK 7. Dendrogram s spodnje slike 29 v bistvu prikazuje grozde, ki se vsebinsko navezujejo na življenjepise in dela tako skladateljev kot tudi književnikov. Prav ključne besede, ki so označile skladatelja ali književnika so mnogokrat dosegle visoke vrednosti K.

Pri hierarhičnem grozdenju pomembnih ključnih besed, ki so že razvrščene v posamezne UDK razrede je postopek grozdenja nekoliko zapletenejši, kajti potrebno je najprej izbrati najbolj značilne kazalce, ki jih je možno med sabo primerjati. V tem pogledu sem se odločil za dejavnik K in števila obiskov po posameznih UDK področjih. Kot prvo sem s pomočjo programskega orodja Excel sešteval vrednosti K za sleherno UDK področje, nakar sem pridobljene vrednosti preračunal v odstotkovne deleže. Prav tako sem preračunal vrednosti o številu obiskov po posameznih UDK področjih v odstotkovne deleže, s čimer sem dobil primerljivost med obema kazalcema. Oba kazalca sta za segmentacijo (*zanimanj*) spletnih obiskovalcev nepogrešljiva. Tako dejavnik »K« poroča o najpomembnejših ključnih besedah z določenega UDK področja, ki so jih uporabili spletni obiskovalci pri poizvedbah na notranjem iskalniku spletnega UDK leksikona. Število obiskov po posameznih UDK področjih mi daje informacijo o tem, za katero področje so se spletni obiskovalci najbolj zanimali. Če združim oba kazalca, dobim povratno informacijo o pogostosti uporabljenih ključnih besed in števila obiskov z določenega UDK področja, na podlagi česar bo možno učinkovito razčleniti ne samo zanimanja spletnih obiskovalcev, temveč bo možno tudi le-te razvrstiti v določene hierarhične grozde. S pomočjo hierarhičnih grozdov je možno ponazoriti razvejenost in oddaljenost med različnimi UDK področij in s tem posledično med spletnimi obiskovalci. Sledi preglednica in slikovni prikaz.

**4.9.6.1 Preglednica 7: Primerjava med dejavnikom najpomembnejših ključnih besed in števila obiskov spletnih obiskovalcev po UDK področjih**

ID	$\Sigma K$	K (%)	$\Sigma No$	No (%)	UDK
1	272,25	5,05	5820	6,2	0
2	264,48	4,91	7479	7,97	1
3	208,16	3,86	9796	10,44	2
4	387,73	7,19	10910	11,63	3
5	499,53	9,27	10237	10,91	5
6	373,29	6,42	8174	8,71	6
7	318,7	5,91	11991	12,78	7
8	389,14	7,22	13830	14,74	8
9	2677,6	49,67	15573	16,6	9
<b>Skupaj</b>	<b>5390,9</b>	<b>100</b>	<b>93810</b>	<b>100</b>	<b>9 razredov</b>



**4.9.6.2 Slikovni prikaz 30: Hierarhično in K – means grozdenje**

Kot prvo bom opisal preglednični prikaz 7, nakar bom v nadaljevanju še opisal slikovni prikaz 30. V preglednici 7 so predstavljeni naslednji kazalci: ID (*vrstni red*), vsota ( $\sum K$ ) K po UDK področjih, odstotkovni delež (%) K, vsota števila obiskov ( $\sum No$ ) po UDK področjih, % delež No in devet razredov UDK, ki so 0, 1, 2, 3, 5, 6, 7, 8 in 9. V spodnjem delu preglednice 7 se nahajajo skupne vrednosti za omenjene kazalce (*za*  $\sum K = 5390,9$ ;  $\%K = 100\%$ ;  $\sum No = 93810$ ;  $\%N = 100$ ). Vrednosti tako za  $\sum K$ ,  $\%K$ ,  $\sum No$  in  $\%No$  po posameznih UDK področjih so naslednji:

$\sum K$  za področje UDK 0 je znašala 272,25 ali 5,05 %,  $\sum No$  je bila 5820, kar odgovarja 6,20 %. Pri področju UDK 1 je bila  $\sum K$  264,48 ali 4,91 %,  $\sum No$  je bila 7479 ali 7,97 %. Pri področju UDK 2 je bila  $\sum K$  208,16 ali 3,86 %,  $\sum No$  je bila 9796 ali 10,44 %.  $\sum K$  za področje UDK 3 je znašala 387,73 ali 7,19 %,  $\sum No$  je bila 10910, kar odgovarja 11,63 %. Pri področju UDK 5 je bila  $\sum K$  499,53 ali 9,27 %,  $\sum No$  je bila 10237 ali 10,91 %.  $\sum K$  za področje UDK 6 je znašala 373,29 ali 6,42 %,  $\sum No$  je bila 8174, kar odgovarja 8,71 %. Pri področju UDK 7 je bila  $\sum K$  318,70 ali 5,91 %,  $\sum No$  je bila 11991 ali 12,78 %. Pri področju UDK 8 je bila  $\sum K$  389,14 ali 7,22 %,  $\sum No$  je bila 13830 ali 14,74 % in nenazadnje za področje UDK 9 je bila  $\sum K$  2677,60 ali 49,67 %,  $\sum No$  je bila 15573 ali 16,60 %.

Zaradi nadaljnjega dela s programskim orodjem Orange Canvas sem s kazalci in ustreznimi vrednostmi (*brez stolpca ID*) s preglednice 7 ustvaril »hierarhija%.txt« datoteko. Desna stran slike 30 prikazuje postopek nalaganja podatkov (*gl. Datoteka*). »Datoteko« sem povezal z »Izbor atributov«, pri čemer sem kot atributa določil  $\%K$  in  $\%No$ , a kot razred (*class*) sem določil UDK. V nadaljevanju sem to povezal z »Razdalja primera (*Example Distance*)«, kjer sem določil Evklidsko razdaljo in kot označevalec UDK. Nenazadnje sem razdaljo primera povezal s hierarhičnim grozdenjem (*povprečna povezanost, anotacija je UDK*). Na takšen način sem kot izid dobil zgornji dendrogram (*gl. sliko 30*). Spodnji dendrogram sem dobil na podlagi zanimivega postopka s K-means grozdenjem. Glavna razlika med hierarhičnim in K-means grozdenjem je v tem, da hierarhično grozdenje samodejno vzpostavlja več grozdov v nekem hierarhičnem razmerju, medtem ko se pri K-means grozdenju ne ukvarjamo s hierarhijami in povrhu tega je potrebno že vnaprej določiti število grozdov. V tem primeru sem določil tri grozde in Evklidsko merilo razdalje. Nadalje sem (*gl. desno stran slike 30*) »K-means« povezal z »Razdalja primera 2 (*evklidska razdalja, označevalec UDK*)«, nakar sem to povezal s »Hierarhično grozdenje 2 (*povprečna povezanost, anotacija z UDK*)«. Z ozirom na to, da se bom v tem delu natančneje ukvarjal z razčlenbo spletnih obiskovalcev glede na njihova (*izobraževalna*) zanimanja in bom tako potreboval hierarhično prikazane podatke, je zame

ključen izid na podlagi hierarhičnega grozdenja. K-means grozdenje sem vpeljal zgolj kot zanimivost, kajti zanimalo me je, če bom kot izid dobil podobno zgradbo in iste razdalje. Povrhu tega me je tudi zanimalo razvrščanje UDK področij v grozde. Kot je možno opaziti z obeh dendrogramov sem dobil kot izid isto zgradbo in iste razdalje, le s to razliko, da vrhnji diagram prikazuje grozde UDK razredov, medtem ko spodnji dendrogram prikazuje tri grozde (tj. 1, 2 in 3), v katere spadajo posamezni UDK razredi. V grozd 1 so se uvrstila naslednja UDK področja: UDK 0, UDK 1 in UDK 2.

V grozdu 2 se nahaja: UDK 6, UDK 5, UDK 3, UDK 7 in UDK 8.

V grozd 3 se je uvrstilo področje UDK 9. Prav slednji zajema vse ostale razrede.

Z gornjega dendrograma na sliki 30 je možno razbrati naslednje:

- grozd UDK 9 zajema vseh devet točk tj. 9, 8, 7, 6, 5, 4, 3, 2, 1 in 0. Višina grozda je enaka 1,137438.

- grozd UDK 8 zajema pet točk tj. 8, 7, 5, 3 in 2. Višina grozda znaša 0,320138. Grozd UDK 8. Točki 3 in 5 sta med sabo povezani in imata drugo najmanjšo razdaljo, ki znaša 0,082792.

- grozd UDK 0 zajema tri točke tj. 0, 1 in 6. Višina grozda je enaka 0,206706. Točki 1 in 6 sta med sabo povezani in imata najmanjšo razdaljo, ki je enaka 0,078418.

Z vidika velikosti grozda UDK 9 bi lahko trdil, da mu je najbolj soroden grozd UDK 8, ki vključuje UDK 7, UDK 5, UDK 3 in UDK 2, nakar šele sledi grozd UDK 0, ki vključuje UDK 1 in UDK 6. Isto zgradbo dendrograma in podobne razdalje dobimo tudi takrat, kadar med sabo primerjamo %N (%številca najmočnejših ključnih besed namesto %K) in %No. Skratka ob primerjavi dveh izbranih kazalcev tj. %K in %No (primerjava med odstotkovnim deležem najmočnejših ključnih besed po UDK področjih in odstotkovnim deležem pogostosti obiskov po UDK področjih) dobim kot izid nekakšno hierarhijo (izobraževalnih) zanimanj spletnih obiskovalcev. Še zlasti zanimivi so izidi % poizvedb, % najmočnejših ključnih besed po izvedeni besedni analizi poizvedb in % pogostosti obiskov za področje UDK 6. Glede pogostosti razvrščenih poizvedb se je področje UDK 6 uvrstilo na drugo mesto, po moči ključnih besed na šesto mesto in glede obiskov na sedmo mesto. To pomeni, da poizvedbe, ki sem jih razvrstil v področje UDK 6 niso toliko odražale dejansko zanimanje spletnih obiskovalcev, kar sem lahko dokazal šele z besedno analizo poizvedb (s tem sem kot izid dobil dejavnike K) in zbranimi spletometričnimi podatki o pogostosti obiskov po posameznih UDK področjih. Prav tako sem ugotovil, da večinoma poizvedbe, ki so bile razvrščene v področje UDK 6 niso bile vsebinsko tesneje povezane s poizvedbami s področja UDK 5 kot sem spočetka predpostavil, iz česar lahko

skleпам, da gre v tem primeru za dokaj nesorodne spletne obiskovalce in poizvedovalce. Zelo presentljiva je bila zame ugotovitev, da so si bile poizvedbe s področja UDK 1 in UDK 6 bolj sorodne. S tem posledično so si tudi bili spletni obiskovalci z obeh omenjenih področij sorazmerno sorodni. Prav ta zame velika ugotovitev je možno obrazložiti na podlagi najmočnejših ključnih besed, kajti poizvedbi kot sta stres in shizofrenija sta se v kategorijo pomembnih ključnih besed z visokim dejavnikom K uvrstili zelo visoko. Prav obe ključni besedi, ki spadata v področje UDK 1, je možno povezovati s pojmi kot so bolezen, bolezenska stanja, zdravje, metode zdravljenja, industrija itd. Prav tovrstne asociacije pa spadajo v področje UDK 6, še zlasti v področje medicine tj. UDK 61 in področje proizvodnje/industrije tj. UDK 65/69 večkrat povezani s ključnimi besedami s področja UDK 6 (*npr. povezave: »stres -> transport, stres -> industrija, stres -> bolezen, shizofrenija -> bolezen, bolezensko stanje, psihologija, shizofrenija -> psihiatrija ipd.«*). Druge asociacije med obema UDK področjema je možno najti npr. za »Etika -> Zdravstvo, Hipokratova zaprisega, Morala -> Zdravniki, Filozofija/psihologija -> Medicina, Temperamenti -> Hipokratovi kvadranti idr.« Z mojega vidika je na podlagi vrhnjega dendrograma s slike 29 tudi presentljivo opažanje, da sta si področje UDK 3 in UDK 5 precej blizu, kajti ugotovljena razdalja med obema področjema je bila (*kot že omenjeno*) zgolj 0,082792. V tem primeru lahko najdemo povezave med npr. ključnim besedam kot so npr. »Ekologija -> sociologija; naravna -> družbena idr.« Še boljše povezave dobimo, če si ogledamo poizvedbe kot npr. Naravoslovje -> družbene vede; Narava -> družba, Statistika kaznivih dejanj -> matematika idr.« V tem primeru gre za spletne obiskovalce, ki se zanimajo tako za družbene kot tudi za naravoslovne pojave z uporabo eksaktnih metod (*npr. statistika družbe -> matematična statistika ipd.*). Najbolj sorodni so si bili spletni obiskovalci s področja UDK 9, UDK 8 in UDK 7, kajti spletni obiskovalci so mnogokrat poizvedovali po življenjepisih slavnih književnikov in skladateljev. Še zlasti zanimivi sta ključni besedi »Življenjepis in Biografija«. Ključna beseda »Življenjepis« je bila sorazmerno pogosta (*frekvenca enaka 39*), vendar je dosegla sorazmerno nizek rang in dejavnik K ( $r = 450$ ;  $K = 3,606$ ). Podobno lahko ugotovimo za ključno besedo »Biografija«, ki je prav tako bila sorazmerno pogosta (*frekvenca enaka 25*) in je dosegla nižji rang ( $r = 89$ ) in nekoliko nižjo vrednost K ( $K = 9,193$ ). Obe ključni besedi sta se večinoma pojavljali ob imenu in priimku znamenitih skladateljev in književnikov, medtem ko so se zgolj imena in priimki le-teh ob poizvedbah pojavljali pogosteje, zaradi česar je so npr. ključne besede kot so Bach, Vivaldi, Seliškar, Aškerc idr. dosegle višji rang in višjo vrednost K. Če si ogledamo razvrstitev v tri grozde zgolj z

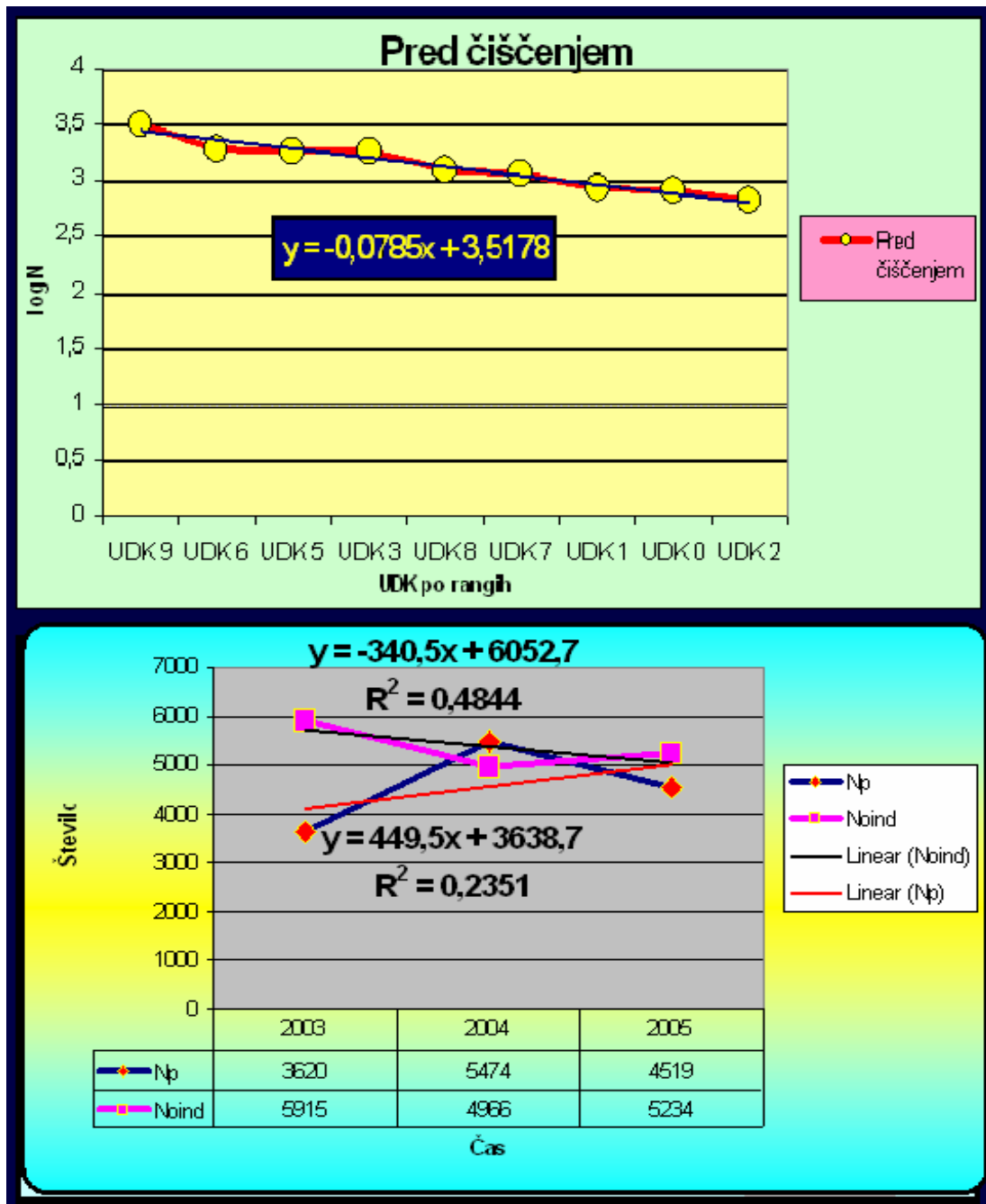
vidika doseženega dejavnika K, potem je možno priti do spoznanja, da spada področje UDK 6 ( $\%K = 6,42 \%$ ) v grozd 2 (gl. *spodnji dendrogram na sliki 30*), medtem ko spada področje UDK 2 v grozd 1 ( $\%K = 3,86 \%$ ).

Na podlagi doslej dobljenih vzorcev oziroma izidov se bom v naslednjem poglavju ukvarjal z ovrednotenjem dobljenih podatkov/izidov (*stabilnost proučevanega sistema, predvidljivost porazdelitev razpona in pomembnost poizvedb, razmerja, kateri podatki naj bi bili na spletni zbirki znanja blizu skupaj ? idr.*) in v nadaljevanju z razčlemba (*izobraževalnih*) zanimanj spletnih obiskovalcev in s tem izpeljal hierarhični miselni model, s katerim je možno segmentirati spletne obiskovalce v različne interesne skupine.

#### **4.9.7 Ovrednotenje podatkov (ugotovitev)/izpeljava novih znanj**

Na podlagi odkritih vzorcev v podatkih bom poskušal le-te najprej ovrednotiti po pomembnosti, nakar bom iz njih izpeljal tista nova znanja, ki naj bi bili za kasnejši miselni model spletnih obiskovalcev pomembni. Kot prvo bi se želel povrniti na prejšnja poglavja, v katerih sem poročal o stabilnosti proučevanega spletnega sistema z ozirom na poizvedovalne vedenjske vzorce spletnih obiskovalcev. Ugotovil sem, da se (*izobraževalna*) zanimanja spletnih obiskovalcev glede na poizvedbe tudi na daljši časovni rok ne spreminjajo. Povrhu tega bi želel osvetliti razmerje med številom poizvedb in številom obiskov na indeksni strani UDK leksikona v časovnem obdobju od leta 2003 do 2006. Zaradi večje nazornosti bom na naslednji strani predstavil ustrezní slikovni prikaz. Zelo pomemben je po moji oceni tudi podatek o načinu glede pogostosti poizvedovanja za nek iskalni pojem, ki poteka po principu hiperbolične odvisnosti (*Zipfova porazdelitev*). Prav ta vedenjski vzorec spletnih obiskovalcev/poizvedovalcev v bistvu poroča o pomembnosti določenih ključnih besed. Na straneh od 88 – 89 sem se na podlagi modela sobe ali poslopja dokopal do pomembne ugotovitve o prvinah (*predmeti oziroma pomembne ključne besede*), ki se najbolj, srednje ali pa najmanj pogosto uporabljajo (*pogoste poizvedbe oziroma pomembne ključne besede*). Velja ponovno omeniti zelo pomemben podatek oziroma pomembno ugotovitev, ki poroča o interdisciplinarni značilnosti močnih ali pomembnih ključnih besedah. Prav na isti strani je možno tudi zaslediti pomembno ugotovitev o ključnih besedah, ki označujejo znamenite osebnosti, še zlasti s področja književnosti, glasbe in svetovne zgodovine, ki so pa v velikem številu primerov posledica programsko usmerjenih poizvedb, še zlasti s strani šolskih programov. Nenazadnje še moram izpostaviti pomembnost izpeljanih asociacijskih pravil in

hierarhičnih grozdov, ki pomenijo nadgradnjo modela (*pomembno za organizacijo podatkov, ki naj bi bili lokacijsko blizu skupaj in za segmentacijo spletnih obiskovalcev*). Za poznejšo segmentacijo spletnih obiskovalcev v interesne skupine je bil ponovno zelo pomemben slikovni prikaz 14 na strani 73 (*okvirna miselna hierarhija*) in v nadaljevanju besedni opis, v katerem sem opozoril na osnovne psihološke vzgibe spletnih obiskovalcev/poizvedovalcev.

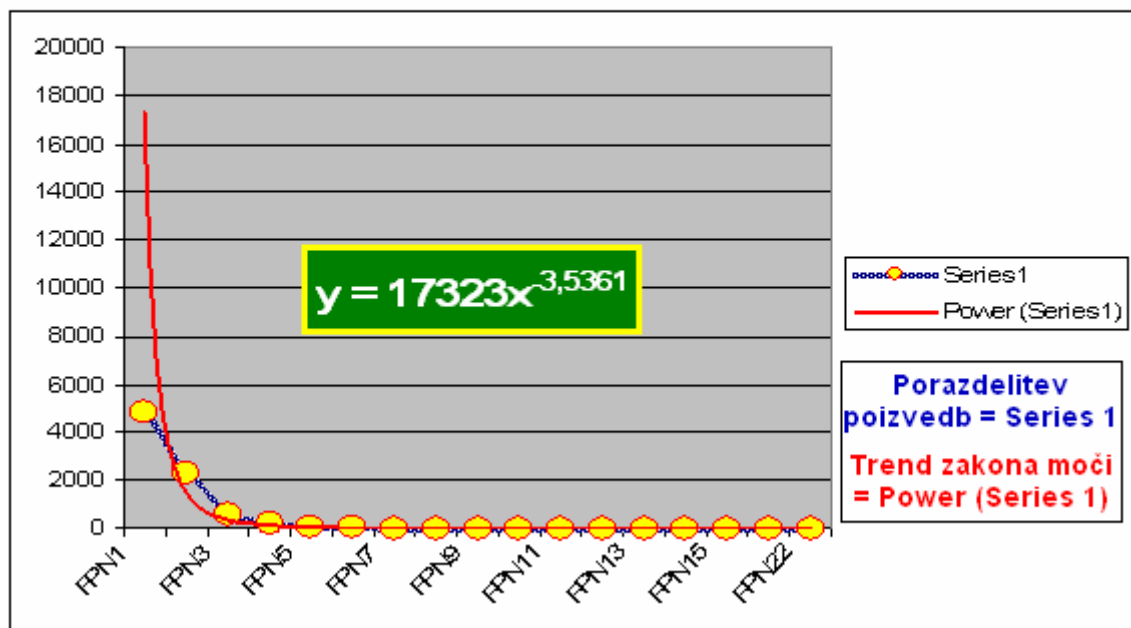


**4.9.7.1 Slikovni prikaz 31: Zakon moči poizvedb, primerjava med številom poizvedb in obiskov ter trendne črte**

Zgornji del slike 31 prikazuje v bistvu isto snov kot levi del slike 13 na strani 69, le s to razliko, da je temu slikovnemu prikazu še dodana (*temno modra*) trendna črta (gl. zapis *linearne funkcije*  $Y = -0,0785 X + 3,5178$ ), ki se tesno prilega rdeči (*skoraj ravni*) črti. Trendna črta dodatno izpostavi dejstvo, da gre pri tem proučevanem pojavu za zakon moči, ki je sorazmerno neodvisen od časa in kjer lahko zapazimo neko močno poudarjeno usmerjenost (*možno je opaziti neko močno vsebinsko usmerjenost tj. še zlasti za področje UDK 9*). Vedenje spletnih obiskovalcev/poizvedovalcev je s tega izhajajoč sorazmerno predvidljivo, iz česar bi lahko sklepal, da imamo v večji meri opraviti s skupinami spletnih obiskovalcev, ki izvedejo poizvedbe na podlagi določenega kolektivnega programa. Tovrstni ljudje so bolj ali manj stalni uporabniki spletnega UDK leksikona. To izpeljano novo znanje o podatkovnem vzorcu dodatno podkrepi spodnji del slike 31.

Spodnji del slike 31 prikazuje primerjavo med številom poizvedb in obiskov skozi leta od 2003 do 2006 in skupaj (gl. *legendo: Np ... število poizvedb – temna črta, linear Np ... trendna črta temne barve; Noind ... število obiskov na indeksni strani – rdeča črta, linear Noind ... trendna črta rdeče barve; gl. tudi preglednico v desnem spodnjem predelu slike 31*). V letu 2003 so spletni obiskovalci izvedli 3620 poizvedb in obiskali indeksno stran 5915 krat (*razmerje med obema je 0,6120*). V naslednjem letu so izvedli 5474 poizvedb in obiskali indeksno stran 4966 krat (*razmerje med obema je 1,1023*). V letu 2005 so izvedli 4519 poizvedb, a število obiskov na indeksni strani je bilo enako 5234 (*razmerje med obema je 0,86340*). V teh treh letih je bilo tako 13613 poizvedb in število obiskov na glavni strani je bilo 16115 (*razmerje, ki pomeni hkrati tudi povprečno razmerje med obema je 0,8447 – isto vrednost dobimo, kadar delimo med sabo povprečno Np s povprečnim Noind tj. 4537,67/5371,67*). Spletni obiskovalci, ki so obiskali indeksno stran UDK leksikona so povprečno v 84,47 % primerih tudi izvedli vsaj eno poizvedbo (*velja za tri leta skupaj*). Trendni črti (*zvezna funkcija za število poizvedb je naraščujoča druga  $y = 449,5x + 3638,7$ , a druga tj. število obiskov na indeksni strani pojema  $y = -340,5x + 6052,7$* ) kažeta na to, da se morda to stanje tudi v prihodnje ne bo bistveno spremenilo. Naslednji vedenjski vzorec spletnih obiskovalcev, ki ga bom v tem poglavju izpostavil je prav tako zanimiv in se dotika pogostosti poizvedovanja za nek določen iskalni pojem. Na naslednji strani bom najprej predstavil slikovni prikaz s preglednico, nakar bom slikovno predstavitev še opremil z besednim opisom. Znaki FPN1..FPN22 predstavljajo kazalci za pogostost poizvedovanja za določen iskalni pojem (X – os), medtem ko bom na Y – os namestil kazalec za skupno število poizvedb.

FPN1	FPN2	FPN3	FPN4	FPN5	FPN6	FPN7	FPN8	FPN9
4865	2320	633	248	84	50	20	14	6
FPN10	FPN11	FPN12	FPN13	FPN14	FPN15	FPN16	FPN22	
8	2	2	1	1	1	1	1	



#### 4.9.7.2 Slikovni prikaz 32: Porazdelitev poizvedb in trend zakona moči

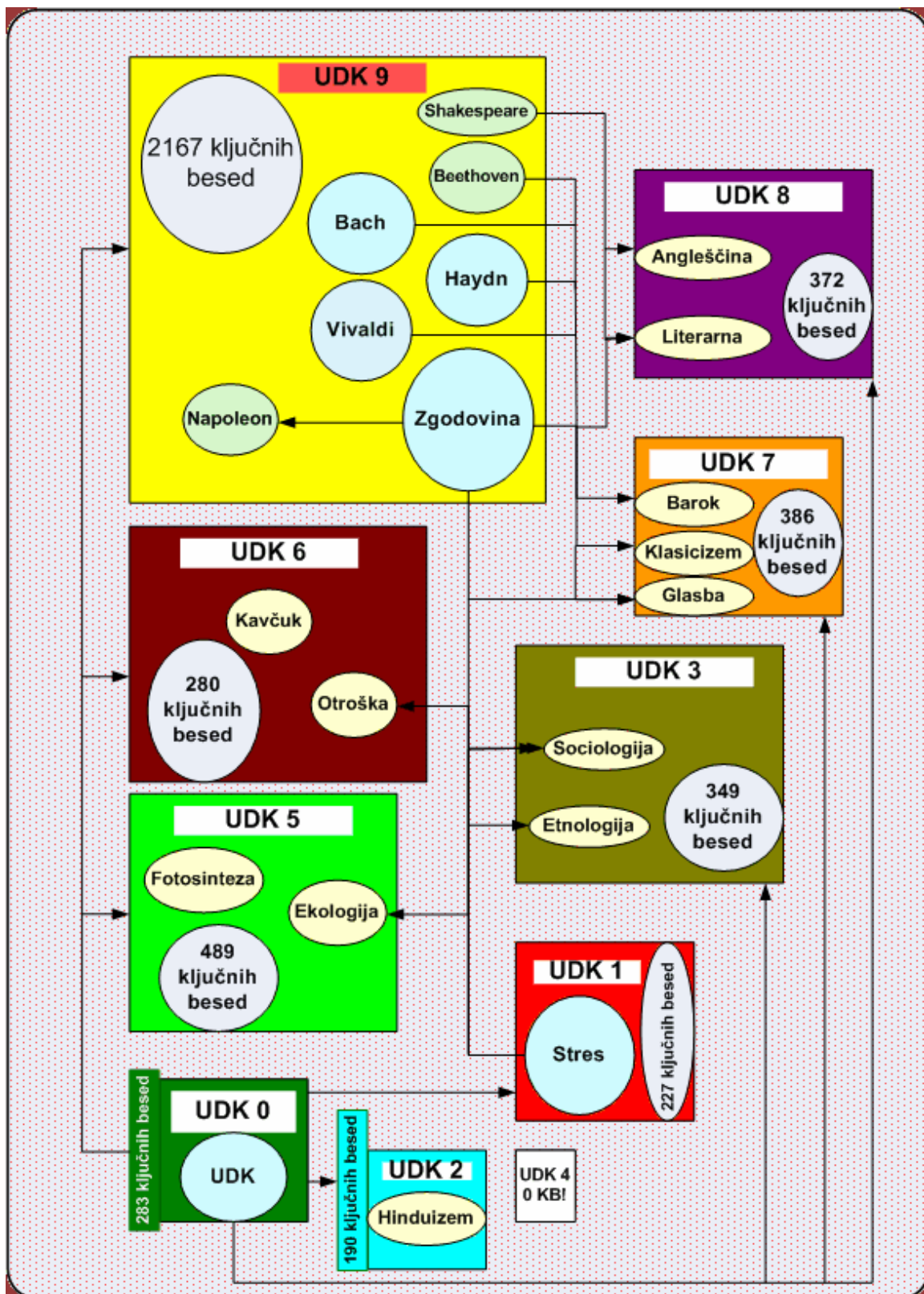
Sika 32 prikazuje porazdelitev poizvedb in trend zakona moči v obliki eksponentne funkcije  $Y = 17323 X^{-3,5361}$  (natančnejši opis porazdelitve poizvedb je dosegljiv na strani 80 - 84 ali pa gl. priloženo preglednico na vrhnjem delu slike 32, zato tega opisa na tem mestu ne bom ponovil). Za proučevane spletne obiskovalce/poizvedovalce je značilen vedenjski vzorec, ki se kakovostno gledano nikoli ne spremeni in je sorazmerno neodvisen od časovne skale, tako da gre tudi v tem primeru za odkrit zakon moči. Spletni obiskovalci bodo v večini primerov za isti iskalni pojem poizvedovali enkrat, dvakrat, trikrat in morda po štirikrat, medtem ko bodo pogostejša poizvedovanja za določen iskalni pojem redkejša (npr. še zlasti poizvedovanje za določen iskalni pojem po 10, 11, 15 in 22 krat je zelo redek pojav). Trendna krivulja zakona moči to ugotovitev še dodatno podkrepi, kajti v prihodnje je možno pričakovati porast 1, 2 in 3 kratnih poizvedovanj za določen iskalni pojem, medtem ko bodo večkratna poizvedovanja relativno gledano ostala tako redek pojav, kot je temu do sedaj bilo primer. Na podlagi zaželene informativnosti in akcijskega udobja, katerim zakonom podlegajo poizvedovalci ali iskalci informacij prav ti nikoli ne bodo spremenili ta vedenjski vzorec, kajti v skladu s ciljem posamezniki kot tudi skupine (v skupinah so združeni posamezniki, ki imajo podobne lastnosti) poskušajo predvideti nepotrebne akcije kot tudi informacijsko vrednost določenega spletnega dokumenta, ki je bil izid poizvedbe na nek določen iskalni pojem.

**Novo znanje/spoznanje:** Ljudje v večini primerov izberejo načine lastnega vedenja, ki je pretežno v skladu s kolektivnimi (*družbenimi*) programi in njihovega pojma o sebi. Prav to znanje/spoznanje je možno prenesti na manjše množice ali skupine ljudi, znotraj katerih imajo posamezniki v sebi določena skladna oziroma podobna subjektivna (*podobne izkušnje*), objektivna (*podobna družbena znanja*) in interpersonalna znanja/spoznanja (*pridobljena znanja/spoznanja s pomočjo empatije in medosebnega sporazumevanja*).<sup>36</sup> Posamezniki in s tem tudi (*homogene*) skupine gledajo na svet okoli sebe skozi prozorne obrazce, ki jih deloma ustvarjajo sami, a jih v večini primerov poskušajo prilagoditi resničnosti, iz katerih se svet sestoji. Po drugi strani pa posamezniki sprejmejo družbene prozorne obrazce, ki jih potem zaradi udobja nekoliko poenostavijo. Obnašanje skupin ima določeno stopnjo generalnosti, ki so v določeni meri tudi hierarhično organizirane in temu ustrezno strukturirane (*zanimanje za določena področja človekovega znanja – npr. interesno področje stresa s poudarkom na kakovostno življenje v procesu dela je hierarhično organizirano na nadrejeni pojem stres, ki so mu podrejeni pojmi kot so zdravje, delovni proces, kakovostno življenje. Prav ta miselna hierarhija predpisuje določen bolj ali manj strukturiran vzorec poizvedovanja kot npr. najprej poizvedovalec izvede poizvedbo »Stres«, nakar oceni kakovost poizvedbe in potem ob ugotovitvi nekakovostnih zadetkov šele oblikuje bolj strukturirano poizvedbo kot npr. »Stres in delovni proces«*). Skratka določeno hierarhično telo (*npr. skupina ljudi*) načeloma vedno išče udobje, informativnost, odprtost nasproti zunanjemu svetu in pozitivno energijo (*tudi v obliki pozitivnih dražljajev iz okolice*), kar je predpogoj za ustrezen pozitiven razvoj oziroma napredek.

S pomočjo modela sobe ali poslopja sem lahko ponazoril površine, ki jih zavzemajo posamezna UDK področja z ozirom na izvedene poizvedbe od leta 2003 do 2006, toda ta predstavitev je zanemarila vsebino teh sob, ki jo lahko opredelim kot pomembne ključne besede znotraj posameznih UDK področij, ki so dosegle visoke vrednosti dejavnika **K**. V ta namen bom na naslednji strani izdelal slikovni prikaz zavzetih površin teh sob v večjem merilu, ki bo prikazal velikosti teh sob in znotraj le-teh v obliki večjih ali manjših obarvanih krogov pomembnost določenih ključnih besed. Povrhu tega bom torvrstne kroge opremil s povezavami do drugih UDK področij ali celo do drugih ključnih besed. Na podlagi prikaza kanim v nadaljevanju izpeljati t.i. interdisciplinarne ključne besede. Povrhu tega bom še izpostavil pomembne ključne besede za zanmenite osebnosti.

---

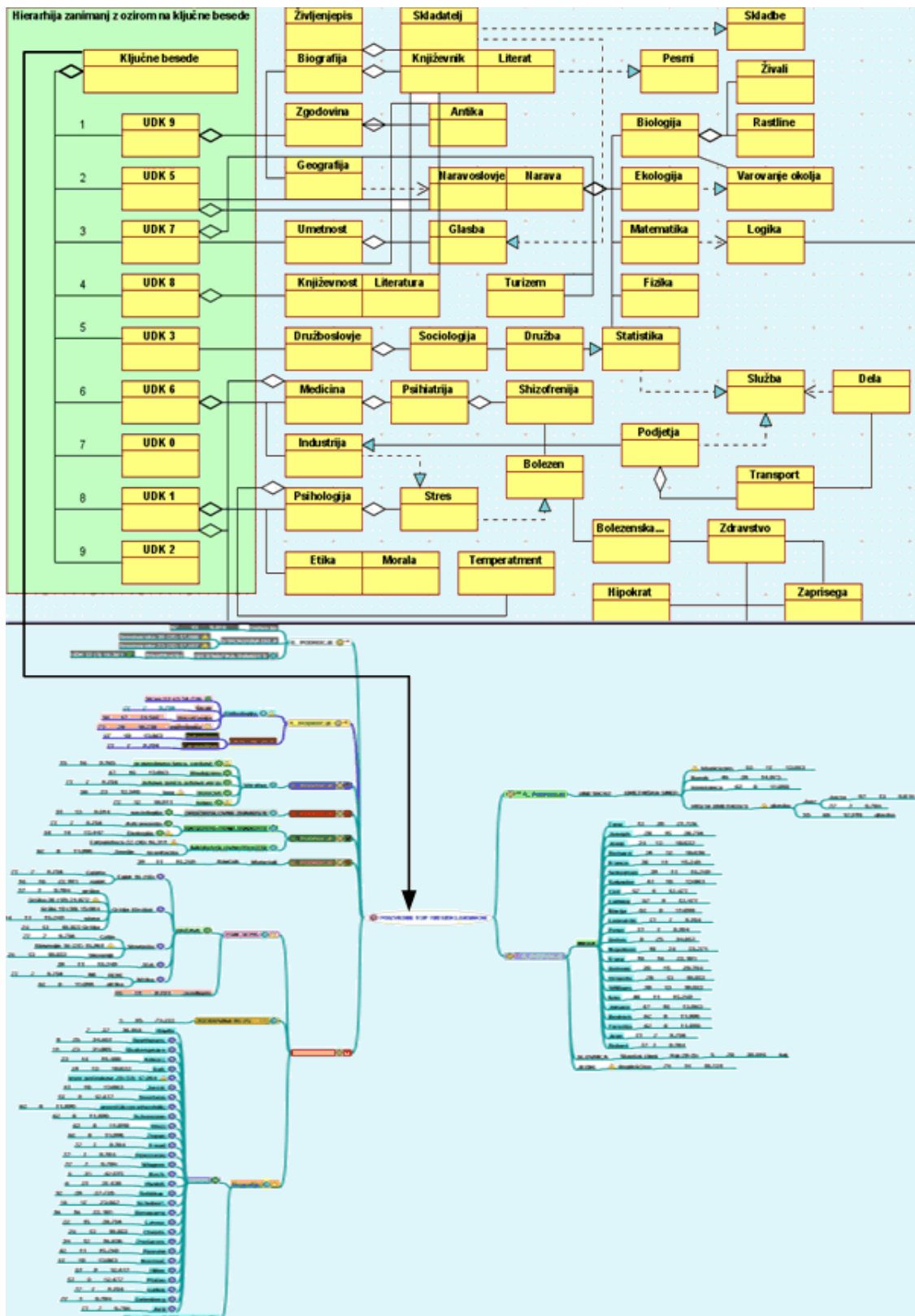
<sup>36</sup> Znanje/spoznanje sem delno izpeljal na podlagi dela: Fulgosi, A.: Psihologija ličnosti; teorija ličnosti na strani 289 (4. odstavek), a delno iz dobljenih izidov.



#### 4.9.8 Slikovni prikaz 33: Model sobe in pomembne ključne besede

Slika 33 prikazuje površinska razmerja med številom poizvedb in pomembnejših ključnih besed s posameznih UDK področij. Povrhu tega so prikazane tudi povezave med posameznimi pomembnejšimi ključnimi besedami in UDK področij. Ocenjujem, da je

slikovni prikaz dovolj zgovoren, tako da s tega vidika natančnejši opis ni potreben. Želel bi pa nekoliko natančneje izpostaviti pomen pomembnejših ključnih besed kot so zgodovina, stres, Vivaldi, Haydn, Beethoven, Bach, Shakespeare in Napoleon, kajti navedne ključne besede imajo z vsebinskega vidika zelo širok spekter. Ključna beseda »Zgodovina (se razvršča pod UDK 9 – gl. velik svetlo moder krog)« je že takšen primer, ki je interdisciplinarne narave, kar pomeni, da jo lahko vsebinsko povežemo z različnimi področij človekovega znanja. Z njenega temelja izhajajoč se navezuje na predmet svetovne in lokalne zgodovine (tj. področje UDK 9). Sicer se lahko ta ključna beseda povezuje z vsemi UDK področij kot (gl. slika 33 in povezave v obliki puščic) npr. **Zgodovina** glasba (znotraj tega barok, klasicizem idr.), literarna **zgodovina**, **zgodovina** ekologija, **zgodovina** sociologija, **zgodovina** etnologija idr. Druga ključna beseda, ki je interdisciplinarne narave je »Stres (se razvršča pod UDK 1 – upodobljena kot večji svetlomoder krog). Ključno besedo »stres« je možno povezati z UDK področij kot so UDK 5 (ekologija), UDK 7 (glasba – zmanjševanje stresa), UDK 6 (otročka – z boleznijo, z otroškim stresom) in z UDK 3 (sociologija – socialni stresni dejavniki). Pri obeh ključnih besedah obstajajo še številne druge povezave, ki jih v tem slikovnem prikazu zaradi prostorskih razlogov nisem uspel umestiti. Druga skupina pomembnih ključnih besed predstavljajo še zlasti priimki znamenitih osebnosti s sveta glasbe, književnosti in svetovne zgodovine. Znamenite osebnosti se razvrščajo pod UDK 9, vendar lahko njihov pomen povežemo tako s področjem UDK 7 (npr. glasba - Beethoven, Vivaldi, Haydn, Bach idr.), s področjem UDK 8 (npr. angleščina in literarna – Shakespeare) kot tudi znotraj področja UDK 9 tj. svetovna zgodovina (npr. svetovna zgodovina – Napoleon). S pomočjo tega vedenja in izpeljanih asociacijskih pravil je možno dognati, kateri podatki na spletni zbirki znanja naj bi se nahajali lokacijsko gledano skupaj. Prav ta del snovi že spada v področje t.i. »pametno« pripravljenih podatkov. Deloma ta problem že razrešuje klasifikacijski sistem UDK, vendar pa le-ta spet ne zagotavlja polihierarhične in poliasociacijske odnose. Tovrsten problem je možno razrešiti z izgradnjo ustreznih semantičnih mrež, ki vsebujejo ustrezne povezave v obliki hierarhij, asociacij, ekvivalentnosti in pojmovne oddaljenosti. Na podlagi izpeljanih asociacijskih pravil in hierarhičnih grozdov je možna izdelava nekakšne okrnjene semantične mreže, kajti za podrobnejšo izdelavo le-te, bi se moral posvetiti pisanju drugega doktorskega dela, ker je omenjena snov izredno zajetna. Prav tega pa v tem delu ni možno že zaradi prostorskih razlogov uresničiti in povrh tega v tej raziskavi zasledujem drug cilj.



**4.9.8.1 Slikovni prikaz 34: Okrnjena semantična mreža hierarhije zanimanj**

Slikovni prikaz 34 ponazarja fragment semantične mreže hierarhije zanimanj spletnih obiskovalcev/poizvedovalcev. Gornji del slike 34 prikazuje hierarhične (črta, ki vsebuje

na koncu bel romb tj. znak za agregacijo – pomeni odnos nadrejenega pojma nasproti podrejenemu), asociacijske (ravna črta – pomeni asociacijski odnos do drugega pojma npr. *Literatura* -> *Književnik ipd.*), generalne (ravna črta, ki vsebuje na koncu trikotnik – označuje nek splošni odnos kot npr. med pojmom *Podjetje in industrijo ipd.*), odvisnostne (prekinjena črta, ki ima na koncu puščico – označuje odnos odvisnosti med npr. pojmom *Službi in delom ipd.*), udejanjanja (prekinjena črta, ki ima na koncu trikotnik – označuje odnos udejanjanja med npr. pojmom *skladatelj in glasbo, kajti skladatelji uresničujejo/ustvarjajo glasbo ipd.*) in pojmovne bližinske odnose (na sliki je ta odnos označen tako, da se dva pojma, ki sta označena kot pravokotnik med sabo spojita npr. odnos pojmovne bližine med pojmom *življenjepis in biografija ali med književnostjo in literaturo ipd.*) med posameznimi vejami človekovega znanja. Spodnji del slike 34 ponazarja miselno mrežo 100 najbolj pomembnih ključnih besed z različnih UDK področij, ki so ovrednoteni na podlagi frekvence, ranga in dejavnika **K**.<sup>37</sup> Ta miselna mreža prikazuje hierarhične odnose med nadrejenimi in podrejenimi pojmi, ki so razvrščeni v UDK skupine v obliki drevesne zgradbe. Glavni namen tega slikovnega prikaza je v tem, da želim izluščiti, katere pojme je potrebno na podlagi izvedenih meritev, opazovanj in analiz organizirati tako, da se bodo v določenem spletnem dokumentu nahajali blizu enega drugega, tako da bodo zlahka dosegljivi spletnemu obiskovalcu. Povrhu tega so tovrstni pripravljene podatki dobra osnova za indeksiranje spletnih dokumentov s pomočjo nadzorne plošče notranjega iskalnika. S slikovnega prikaza 34 bom kot zgled vzel izhodiščni pojem »Biografija = Življenjepis (se razvršča z globalnega vidika v področje UDK 9) in ustvaril nekakšno pojmovno verigo. **Izpeljano novo znanje/spoznanje:**

1. hierarhični nivo: Biografija = Življenjepis (*sinonima*): asociacija Osebe, Rodbine
2. hierarhični nivo: Skladatelj = Kompozitor; Slikar; Arhitekt; Kipar; Književnik = Literat; Vladar; Politik; Znanstvenik – Naravoslovec – Družboslovec; Aplikativci : asociacija Življenje in izdelki/stvaritve
3. hierarhični nivo: življenjska pot - glasbena dela (*sonate, opere, fuge, madrigali itd.*) - slikarska dela (*vodenke = akvareli, plastenke, oljnata platna idr.*), kiparska dela (*glina, železo, kamen idr.*) - arhitekturne stvaritve (*Eiflov stolp, First Leiter Building, Katedrala idr.* – književna dela, poezija, proza, literarna znanost, literarna zgodovina, literarna kritika, literarno izobraževanje idr.) – izumi (*teleskop, mikroskop idr.*) – odkritja (*Vasco Da Gamma, Krištof Kolumb idr.*) – inovacije (*organizacijske, tehnološke idr.*)

<sup>37</sup> V spremnem gradivu tega dela na CD-ROM si bo možno takšno mrežo поблиže ogledati!

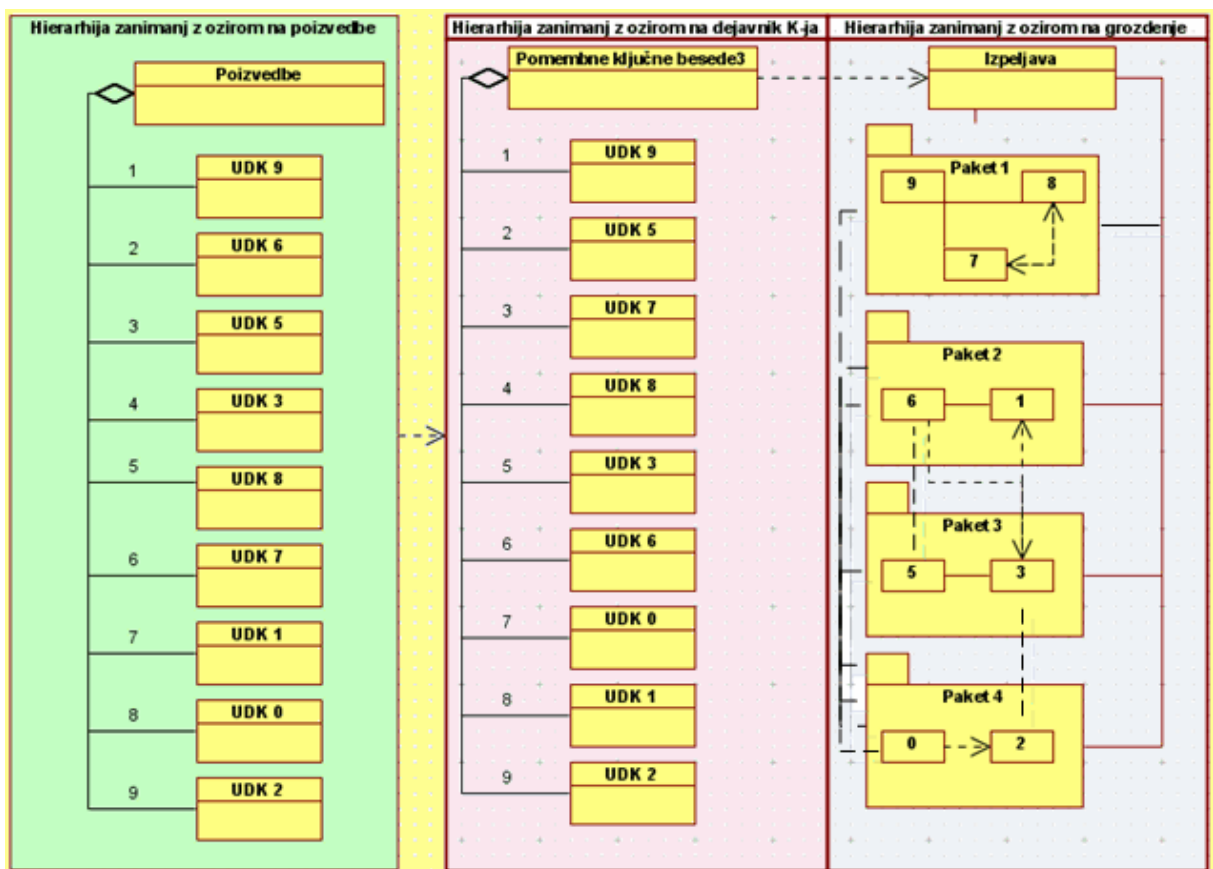
4. hierarhični nivo je multidisciplinarne in interdisciplinarne narave, ki omogoča polihierarhične, poliekvivalentne in poliasociacijske odnose kot npr. povezava med Osebo/skladatelj - dela/ izdelki glasbena dela -> zdravljenje stresa s pomočjo glasbe -> pozitiven vpliv glasbe skladatelja Georga Gershwina na delovno vzdušje v podjetju -> računalniška glasba -> filozofija življenja Johanna Sebastian Bacha in Wolfgang Amadeusa Mozarta -> cerkvena glasba in cerkveni obredi -> Richard Wagner in ples itd. Zgled »Stres«:

1. hierarhični nivo: UDK 1 – Psihologija : asociacija posameznik /-i, osebnost /-i
2. hierarhični nivo: stres – teorija - klasifikacija – stresni dejavniki = stresorji - evstres – distress : pozitivni vplivi/negativni vplivi – znanstveniki -> proučevanje stresa
3. hierarhični nivo: zdravljenje stresa, zdravljenje s stresom
4. hierarhični nivo je prav tako multidisciplinarne in interdisciplinarne narave kot pri predhodnem zgledu: npr. stres -> družba -> stres kot družbeni problem -> gospodarstvo -> svetovni problem stresa -> statistika -> stres po državah Evropske skupnosti in ostalega sveta, okolje -> stres -> varstvo okolja, delovno mesto -> stresni dejavniki, stres -> turizem, stres -> duševne motnje -> zdravljenje shizofrenije, stres -> zdravila proti stresu, stres -> fizikalna odpornost materialov, stres -> entropija -> kemijske reakcije, stres -> proizvodnja -> transport, stres -> raztezna napetost gradbenih delov, stres -> promet -> vozila -> vozniki -> alkohol – droge - narkoleptiki idr.

Za ustvarjanje učinkovite pojmovne mreže, ki zmore prikazati različne odnose med pojmi potrebujemo kot prvo nek globalni miselni model, ki zajema tako različne spremenljivke kot tudi posebnosti le-teh. V nadaljnjem razvoju tovrstne pojmovne mreže je potrebno le-takšno redno dopolnjevati z ustreznimi podatki. Prikaz tovrstnih pojmovnih mrež bi moral biti v obliki drevesnega slikovnega prikaza in/ali besednega opisa, ki bi omogočal vrtanja navzgor po hierarhiji kot tudi vrtanja navzdol po hierarhiji pojmov in hkrati krmiljenje po sorodnih in/ali pojmihi ki so v ožji povezavi z izhodiščnim pojmom. Povrhu tega bi moralo takšno pojmovno drevo tudi omogočati krmiljenja po drugih področjih znanosti, umetnosti, aplikacij in športa, znotraj katerih bi bilo zopet možno vrtanje navzgor in navzdol po hierarhiji. **Tudi v svetu znanja obstajajo določene generalnosti oziroma določeni miselni principi, ki so prav tako hierarhično organizirani in strukturirani, saj izhajajo iz človeka samega!** V naslednjem podpoglavju 4.9.9 se bom ukvarjal s segmentacijo spletnih obiskovalcev/poizvedovalcev v interesne skupine.

### 4.9.9 Segmentacija spletnih obiskovalcev/poizvedovalcev

Na strani 73 (gl. slika 14) sem že izpostavil okvirno miselno hierarhijo (izobraževalnih) interesov spletnih obiskovalcev/poizvedovalcev, nakar sem s pomočjo metode asociacijskih pravil in hierarhičnega grozdenja izluščil podrobnosti. Še zlasti metoda hierarhičnega grozdenja mi je dajala natančnejše podatke, kajti ob upoštevanju dejavnika K in števila obiskov po UDK področjih sem dobil kot izid hierarhično zgradbo z ozirom na evklidske razdalje oziroma bližine med obravnavanimi področji. Naj v zvezi z zapisanim predstavim pregledno shemo dobljenih izidov za poizvedbe, pomembne ključne besede in hierarhičnih grozdov.



#### 4.9.9.1 Slikovni prikaz 35: Pregled (izobraževalnih) zanimanj spletnih obiskovalcev/poizvedovalcev

Slika 35 ponazarja tri stolpce hierarhij (izobraževalnih) zanimanj spletnih obiskovalcev/poizvedovalcev z ozirom na analizo poizvedb, besedni analizi poizvedb in izpeljavi na podlagi asociacijskih pravil in hierarhičnega grozdenja. Številke od 1 do 9 pomenijo dosežen rang, ki so jih posamezna UDK področja dosegle ali pri analizi ali pa pri besedni analizi poizvedb. O tem sem že poročal v prejšnjih poglavjih, zato bi bilo nesmotrno to na tem mestu ponoviti. Nekoliko podrobneje bom opisal tretji stolpec s

slikovnega prikaza 35. Pri tem gre za izpeljavo hierarhije (*izobraževalnih*) zanimanj spletnih obiskovalcev/poizvedovalcev na podlagi asociacijskih pravil in še zlasti hierarhičnih grozdov. Sestavil sem štiri pakete, v katerih se nahajajo oštevilčena območja (npr. 6 – *področje UDK 6*), ki so med sabo različno močno povezana. V paketu 1 se nahajajo UDK področja 9, 8 in 7. Področje UDK 9 je s področjema UDK 8 in UDK 7 močno povezano (*gl. ravni črti*), medtem ko obstaja med področjema UDK 8 in UDK 7 obojestranska šibka povezava (*gl. obojestransko črtkano povezavo s puščicama*). Pomenska interpretacija paketa 1 je naslednja (*o tem sem sicer že poročal na prejšnjih straneh od tega dela, a zaradi izpeljave pomena ostalih paketov sem navedeno zaradi večje nazornosti ponovil*):

a.) Za področje UDK 9

Razvoj sveta in človeka. Gibanje Zemlje, pojavi, vreme, pomembni zgodovinski dogodki na svetovnih in lokalnih ravneh, znamenite osebnosti in družine, ki so (so)vplivale na razvoj družbe in s tem morda posredno na sam razvoj Zemlje.

b.) Za področje UDK 8

Jezik ima pravila in svojo zgradbo in pomeni sredstvo za različne načine sporazumevanja. Sporočanje poteka slušno in pisno. V književnosti se opisuje življenje fiktivno in/ali resnično.

c.) Za področje UDK 7

Čutna izrazila in estetika soblikujejo življenje oziroma kulturo bivanja. Poteka način komunikacije med izvajalci in poslušalci, ki je lahko slušen, tipen, vizualen ali gibalen. Izredno pomemben je ritem, ki bi lahko primerjali z ritmom narave in/ali življenja.

Ob upoštevanju globalnih logičnih značilnosti UDK področij in dobljenega izida sem izpeljal novo semantično trditev, ki je naslednja: »**Krajšanje prostega časa, zabava, spoznavanje novih ali manj vsakdanjih stvari. Povpraševanje po izjemnih stvareh, krajev in ljudi.**« **Opomba: V večini primerov so bile poizvedbe s teh področij programsko spodbujene s strani šolskih ali študijskih programov.**

Paket 1 je v šibki pomenski povezavi z ostalimi paketi (*gl. črtkane povezave*). Zgolj področje UDK 9 v paketu 1 se lahko posredno navezuje z ostalimi paketi in s tem posledično s posameznimi UDK področij takrat, kadar gre za poizvedbe v zvezi z življenjepisi slavnih družboslovnih in/ali naravoslovnih znanstvenikov, verskih poglavarjev ter politikov (*UDK 1, UDK 2, UDK 3, UDK 5 in UDK 6*), o slovarjih, leksikonov, enciklopedij ipd. (*UDK 0*) in geografija (*geografija se sicer razvršča pod UDK 9, vendar se lahko povezuje s področjem npr. napovedovanje vremena, ki se*

razvršča pod UDK 5 ali pa poimenovana država se razvršča v področje UDK 9, medtem ko se država kot pravno, gospodarsko in politično telo razvršča pod UDK 3). Zdaj bom poročal o glavnih značilnostih paketa 2, nakar bom v nadaljevanju izpeljal novo semantično trditve. V paketu 2 je možno opaziti, da obstaja močna povezava med področjem UDK 6 in UDK 1 (gl. *ravno črto na sliki 35*), obstaja šibka povezava med UDK 6 in UDK 5 (gl. *črtkano črto*), obstaja enosmerna srednje šibka povezava med UDK 6 in UDK 3 (gl. *črtkano črto in puščico*) in obstaja dvosmerna šibka povezava med UDK 1 in UDK 3 (gl. *črtkano črto in dve puščici*). Paket 2 je v zelo šibki povezavi s paketom 1 in paketom 4 (gl. *črtkane črte*).

Izpeljava nove semantične trditve:

Za področje UDK 6 - gre za bolezni, zdravljenja, zdravilne pripomočke, anatomijo človeka, zdravstvene metode, stroje, električna omrežja, gradbena dela, industrije itd.

Za področje UDK 1 - se osredotoča na bistvo in pomen človeka, še zlasti s poudarkom na posamezniku, ki je zahvaljujoče svojemu psihološkemu ustroju (*miselni procesi, čustvovanje, spominjanje ipd.*) v nenehnem stiku z naravo in z družbo. Prav ta stik pa človeku neobhodno predpisuje nek logičen in kakovosten odnos do okolja in do samega sebe (*etika, morala ipd.*). Skratka pomen področja UDK 1 se lahko strne na označevanje določenih človeških lastnosti in/ali logičnih postopkov/ravnanj.

Nova (*združena*) semantična trditev je lahko naslednja:

**Obstoj (*ohranitev*) človeka temelji na osnovi psihičnega in fizičnega zdravja (*bolezen pomeni opozorilni znak in vzpostavljanje ustreznega ravnotežja*), kar je predpogoj za učinkovito in kakovostno delovanje posameznika v združbah tj. razreševanje oprijemljivih problemov, izdelava uporabnih stvari, ustrezna filozofija življenja ipd.**

V paketu 3 obstaja močna povezava med področjema UDK 5 in UDK 3 (gl. *ravno črto*).

Ob poprej ugotovljenih šibkih in srednje šibkih povezavah lahko navedem še zelo šibko povezavo med področjema UDK 3 in UDK 2 (gl. *črtkano črto*). Ob postopku izpeljave nove semantične trditve je potrebno najprej navesti globalni pomen obeh področij v strnjeni obliki:

Za področje UDK 5

Vso našo izvorno naravno okolje, ki ga kot takšnega prepoznamo in priznamo ter s katerim smo v nenehnem stiku, se večinoma nahaja izven našega človeškega telesa (*izjeme so lahko npr. virusi, bakterije, paraziti - mikrosvet*). Ta svet učinkuje oziroma vpliva na nas ljudi v zelo veliki meri, čeprav vrši človeška vrsta v vsej svoji evoluciji precejšen vpliv na naravo, ki pa ni vedno bil zelo etičen in moralen. Zbirajo se v veliki meri teoretična

spoznanja in z eksperimenti pridobljeni dokazi, ki se v nadaljnji stopnji praktično uporabljajo.

Za področje UDK 3

Ko je vsebina človeka osmišljena z bistvom, etiko, moralo, logiko, z višjimi miselnimi procesi in s pravo pozitivno vero in/ali ideologijo, se lahko posamezniki med seboj povezujejo v družbo /-e. To je vse, kar je z vidika človekovega obstoja in dejavnosti družbenega predznaka (*medosebno sporazumevanje, družbena pravila, izobraževanje, prebivalstvo, narodi, mednarodne konvencije ipd.*).

Nova (*združena*) semantična trditev je naslednja:

**Družbeno in naravno okolje se ovrednotita na podlagi družbenih sporazumov, ki se raztezajo od pravilnikov, zakonov pa tja do računskega določevanja tega sveta. Skrb za čistost in lepote družbenega ter naravnega okolja je ena od glavnih poslanstev človeške vrste in s tem posledično tudi pogoj za preživetje, ki ga omogoča med drugim tudi tradicija.**

Paket 4 pomeni z ozirom na dobljene izide najšibkejša združba, kjer lahko zapazimo enosmerno šibko povezavo področja UDK 0 s področjem UDK 2 (*gl. prekinjeno črto s puščico*). Po drugi strani je ta paket povezan, čeprav po moči zelo šibko, z vsemi ostalimi paketi oziroma UDK področij, kajti področje UDK 0 pomeni človekovo znanje, ki je ali zelo splošne (*npr. leksikoni, slovarji, enciklopedije*) ali pa interdisciplinarne narave (*npr. informatika, bibliotekarstvo, sistematika znanosti*). V bistvu gre pri tem področju za osnovna znanja, ki naj bi jih današnji splošno razgledani civilizirani človek posedoval. Pri področju UDK 2 gre v bistvu za neko široko in globoko podoživljanje sveta, ki ga poskušamo s pomočjo znanja in čustev dojeti. Z ozirom na paket 4 bi lahko izpeljal naslednjo semantično trditev:

**Želja človeka, da bi bil vseveden, s široko piramido znanja, ki je dodatno podkrepljeno z vero v silo/-e, ki ščitijo človekov obstoj (*želja po nesmrtnosti, vsemogočnosti, strah pred padanjem idr.*).**

S pomočjo predhodne razčlenbe hierarhije (*izobraževalnih*) zanimanj bom lahko s pomočjo primerjave med semantičnimi trditvami in globalnimi miselnimi silnicami spletnih obiskovalcev/poizvedovalcev izpeljal model segmentacije spletnih obiskovalcev v posamezne skupine. V ta namen bom najprej izdelal matriko moči in oddaljenosti povezav med UDK področij psiholoških vzgibov.<sup>38</sup> Preden bom sestavil ustrezno matriko naj najprej navedem

---

<sup>38</sup> Glede psiholoških vzgibov sem se vzgledoval po naslednjem delu: Sudar J.(1971): Ekonomska propaganda. 3. dopunjeno izdanje. - Zagreb: Informator.

osnovne oziroma globalne miselne značilnosti človeka (*spletnega obiskovalca*). Za učinkovito segmentacijo spletnih obiskovalcev je po mojem mnenju potrebno poznati globalne miselne silnice posameznikov, znotraj katerih so vsebovani psihološki vzgibi.

Sleherni človek razmišlja bolj ali manj intenzivno na naslednje načine:

- a.) Vsakdanji način razmišljanja, ki vključuje splošne kvalitete in kvantitete kot so dobro počutje, hrana, počitnice, družabnost, zabava, igra, itd.
- b.) Filozofski način razmišljanja, ki vključuje znanost, umetnost, poslovanje, inovativnost itd
- c.) Libidni način razmišljanja, ki vključuje seks, ljubezen, sovraštvo, čustva itd.

S pomočjo tega globalnega miselnega sistema in izpeljanih semantičnih trditvev po paketih bom segmentiral spletne obiskovalce v skupine.

V tej piramidi mišljenja, v kateri se prepletajo različni psihološki vzgibi, se manifestirajo različni cilji in različna dejanja ljudi v civiliziranem svetu. Na podlagi istovetnih ciljev in podobnih dejanj različnih ljudi, je možno le-te razvrstiti v skupine, kar precej olajša njihovo razčlenbo. Osnovne psihološke vzgibe ljudi je možno spodbuditi, ohraniti ali pa zavirati.

Osnovni psihološki vzgibi pri spletnih obiskovalcih so naslednji:<sup>39</sup>

- 1.) Tek po hrani in pijači - hrano in pijačo je možno povezati s področjem UDK 6, v katero se razvršča tudi živilska industrija, živila ipd. Spada pod vsakdanji način razmišljanja!
- 2.) Ljubezen k nasprotnemu spolu (*te je možno povezati s področjem UDK 1* tj. s čustvi *oziroma psihologijo*). Spada pod libidni način razmišljanja.
- 3.) Družinska ljubezen (*povezanost s področjem UDK 3, družina je najmanjša prvina družbe*). Spada pod libidni način razmišljanja.
- 4.) Udobje – možna povezava s področjem UDK 6, UDK 1, UDK 3, UDK 7, UDK 8, UDK 9 in UDK 0. Spada pod vsakdanji način razmišljanja!
- 5.) Slo za uspehom – možna povezava z UDK 1, UDK 3, UDK 9 – npr. slavni ljudje. Spada pod filozofski način razmišljanja.
- 6.) Strah – možna povezava s področjem UDK 1, UDK 2, UDK 3, UDK 9. Spada pod libidni način razmišljanja!
- 7.) Slo po vedoželjnosti – možna povezava s področjem UDK 0 - npr. informatika - programska oprema, UDK 6 – npr. Avtomobili ipd. Spada pod filozofski način razmišljanja.

---

34 Izpeljal iz seminarske naloge z naslovom »Ekonomski propagandni program v službi splošnoizobraževalnih knjižnic : spodbujanje osnovnih psiholoških vzgibov uporabnikov, ki sem jo izdelal leta 1998.

- 8.) Konstrukcija (*npr. gradnja hiše*), kar se da povezati z družinsko ljubezenijo, z UDK 6 – gradbeništvom, z avtomobilsko industrijo, z UDK 7 – arhitekturo, umetnosti, z UDK 5 npr. naravni procesi, naravno okolje, razvijanje teorij ipd. Spada pod filozofski način razmišljanja.
- 9.) Želja po potovanju in gibanju – možna povezava s področjem UDK 9 – npr. letovanje na Kreti itd. in UDK 7 – npr. Šport, turizem ipd. Spada pod vsakdanji način razmišljanja.
- 10.) Zdravje – možna povezava s področjem UDK 1, UDK 3, UDK 5 in UDK 6. Spada pod vsakdanji način razmišljanja.
- 11.) Družabnost – še zlasti povezano s področjem UDK 1, UDK 2 – npr. cerkveni obredi UDK 3, UDK 7 – npr. družabne igre ipd., UDK 8 – večerno branje v manjši družbi ipd., UDK 9 – skupinska popotništva, genealoška društva. Spada pod vsakdanji način razmišljanja.
- 12.) Sožalje (sočustvovanje) – povezljivo z UDK 1, UDK 2, UDK 3, UDK 7. Spada pod libidni način razmišljanja.
- 13.) Tekmovalnost – možna povezava z UDK 0, UDK 1, UDK 2, UDK 3, UDK 5, UDK 6, UDK 7, UDK 8 in UDK 9. Spada pod filozofski način razmišljanja.
- 14.) Okraševanje (*estetika, lepota*) – možna povezava z UDK 0, UDK 1, UDK 2, UDK 3, UDK 5, UDK 6, UDK 7, UDK 8 in UDK 9. Spada pod filozofski način razmišljanja.
- 15.) Šport (igra) - možna povezava z UDK 0, UDK 1, UDK 2, UDK 3, UDK 5, UDK 6, UDK 7, UDK 8 in UDK 9. Spada pod vsakdanji način razmišljanja.
- 16.) Lov (*eksistenca, pustolovščina, kariera, znanje, šport ipd.*) - možna povezava z UDK 0, UDK 1, UDK 2, UDK 3, UDK 5, UDK 6, UDK 7, UDK 8 in UDK 9. Spada pod filozofski način razmišljanja.
- 17.) Harmonija (*lepota*) - možna povezava z UDK 0, UDK 1, UDK 2, UDK 3, UDK 5, UDK 6, UDK 7, UDK 8 in UDK 9. Spada pod filozofski način razmišljanja.
- 18.) Humor - možna povezava z UDK 0, UDK 1, UDK 2, UDK 3, UDK 5, UDK 6, UDK 7, UDK 8 in UDK 9. Spada pod vsakdanji način razmišljanja.
- 19.) Čistoča - možna povezava z UDK 0, UDK 1, UDK 2, UDK 3, UDK 5, UDK 6, UDK 7, UDK 8 in UDK 9. Spada pod filozofski način razmišljanja.
- 20.) Radovednost - možna povezava z UDK 0, UDK 1, UDK 2, UDK 3, UDK 5, UDK 6, UDK 7, UDK 8 in UDK 9. Spada pod filozofski način razmišljanja.
- 21.) Lojalnost/zvestoba - možna povezava z UDK 0, UDK 1, UDK 2, UDK 3, UDK 5, UDK 6, UDK 7, UDK 8 in UDK 9. Spada pod vsakdanji način razmišljanja.

Vse navedene vzgibe je možno povezati z različnimi UDK področij, vendar so te povezave različno močne ali šibke. **Glavni namen primerjave oziroma ovrednotenja paketov in globalnih miselnih silnic, v katerih se nahajajo navedeni psihološki vzgibi je v tem, da**

**želim odkriti/razkriti temeljne miselne hierarhije spletnih obiskovalcev, ki so v obdobju od 2003 do 2006 izvedli poizvedbe na notranjem iskalniku UDK leksikona, tako da jih bo možno kakovostno razvrstiti v skupine.** Ocenjujem, da lahko na podlagi poznavanja globalnih miselnih značilnosti (*posredno psiholoških vzgibov*) spletnih obiskovalcev lažje ovrednotimo njihova (*izobraževalna*) zanimanja oziroma morda lahko razkrijemo osnovne silnice, zaradi katerih se zanimajo za določeno/-e vsebino/-e (*npr. skupina spletnih obiskovalcev se zanima za področje športa, iz česar bi lahko sklepali, da je možno njihovo zanimanje še dodatno spodbuditi z dražljaji, ki spodbudijo psihološki vzgib po tekmovalnosti, po družabnosti, pustolovščini, zdravju ipd.*). Po drugi strani pa bi morda lahko bilo lažje odgovoriti na vprašanje zakaj se tovrstni ljudje zanimajo za določeno področje človekovega znanja (*če je določen segment globalne miselnosti in s tem posredno določen psihološki vzgib pri eni skupini spletnih obiskovalcev bolj izrazit kot pri drugi skupini, bi lahko lažje napovedali njihova dejanja, njihove potrebe in tako izboljšali tako vzgojnoizobraževalne storitve kot tudi povečali npr. prodajo določenih proizvodov, v primeru osredotočenosti na poslovanje idr.*)? Za izdelavo modela adaptivne digitalne knjižnice je po moji oceni potrebno poznati različne skupine spletnih obiskovalcev (*to velja tudi za poznejšo organizacijo informacij kot tudi za izdelavo miselnega modela tj. za aplikacijo adaptivne digitalne knjižnice*). Pri izdelavi matrike moči poudarka med glavnima entitetama „Globalne miselne silnice (*podenote so psihološki vzgibi*) in Paketi (*paketi so opredeljeni z že izpeljanimi semantičnimi trditvami s posameznih UDK področij*)“, se bom naslanjal na svoje magistrsko delo iz leta 2005.<sup>40</sup> Moči poudarkov bom ovrednotil na številčni lestvici z vrednostmi od 0 do 3, pri čemer vrednost 0 opredeli zelo šibak poudarek, 1 – šibak poudarek, 2 – srednje močan poudarek in nenazadnje vrednost 3 opredeli močan poudarek. Skratka primerjati nameravam med sabo izpeljane semantične trditve z globalno miselno podstavo, ki vsebujejo oziroma zajemajo osnovne psihološke vzgibe kot podenote. Na ta način bom na eni strani dobil kot izid moč poudarka določene globalne miselne podstave (*posredno dobim moč poudarka psiholoških vzgibov*) in po drugi strani bom dobil vpogled v moč paketov (*posredno dobim moč poudarka UDK področij, ki se nahajajo kot podenote znotraj teh paketov*). Sledi matrika v obliki sestavljene preglednice.

---

<sup>40</sup> Zamisel sem izpeljal na podlagi branja naslednjega dela: Vester, F.(1991). Kriza prenaseljenih območji : o razvijanju ekosistemskega mišljenja. Ljubljana: DZS, na strani od 118 - 119

#### 4.9.9.2 Preglednica 8: Matrika segmentacije spletnih obiskovalcev v skupine

Pomen znakov	Hierarhija zanimanj				
0 ... zelo šibak poudarek	Paket 1	Paket 2	Paket 3	Paket 4	OMSP
1 ... šibak poudarek					
2 ... srednje močan poudarek					
3 ... močan poudarek					
OMSP ... vsota osnovnih miselnih silnic z ozirom na pakete					
Osnovne miselne silnice	UDK področja				VG
Vsakdanji način razmišljanja	3	3	2	2	10
Filozofski način razmišljanja	3	3	3	3	12
Libidni način razmišljanja	2	3	1	0	6
<b>Vsota po paketih</b>	<b>8</b>	<b>9</b>	<b>6</b>	<b>5</b>	<b>28</b>
<b>Število obiskov</b>	41394	15653	21147	15616	93810
<b>Število poizvedb</b>	5538	2763	3689	1469	13459
<b>Dejavnik K</b>	3385,44	637,77	887,26	480,41	5390,88

**Semantične trditve**

**Paket 1 (UDK 9, UDK 8, UDK 7):**  
»Krajšanje prostega časa, zabava, spoznavanje novih ali manj vsakdanjih stvari. Povpraševanje po izjemnih stvareh, krajev in ljudi.«

**Paket 2 (UDK 6, UDK 1):**  
Obstoj (ohranitev) človeka temelji na osnovi psihičnega in fizičnega zdravja (bolezen pomeni opozorilni znak in vzpostavljanje ustreznega ravnotežja), kar je predpogoj za učinkovito in kakovostno delovanje posameznika v združbah tj. razreševanje oprijemljivih problemov, izdelava uporabnih stvari, ustreznost filozofija življenja ipd.

**Paket 3 (UDK 5, UDK 3):**  
Družbeno in naravno okolje se ovrednoti na podlagi družbenih sporazumov, ki se raztezajo od pravilnikov, zakonov pa tja do računskega določevanja tega sveta. Skrb za čistost in lepote družbenega ter naravnega okolja je ena od glavnih poslanstev človeške vrste in s tem posledično tudi pogoj za preživetje, ki ga omogoča med drugim tudi tradicija.

**Paket 4 (UDK 0, UDK 2):**  
Želja človeka, da bi bil vseveden, s široko piramido znanja, ki je dodatno podkrepjeno z vero v silo/-e, ki ščitijo človekov obstoj (želja po nesmrtnosti, vsemogočnosti, strah pred padanjem idr.).

Opomba: V večini primerov so bile poizvedbe s teh področij programsko spodbujene s strani šolskih ali študijskih programov.

Preglednica 8 ponazarja matriko segmentacije spletnih obiskovalcev/poizvedovalcev v skupine z ozirom na njihove značilnosti (*izmerjene in ovrednotene*), ki so naslednje:

- Osnovne miselne silnice, ki se členijo na vsakdanji način razmišljanja, filozofski način razmišljanja in libidni način razmišljanja.
- Hierarhija zanimanj (*gl. pakete in semantične trditve*), ki se členi na pakete 1, 2, 3 in 4 in posredno na skupine UDK področij 9,8,7,6,1,5,3,0 in 2.
- Število obiskov po posameznih UDK področjih z ozirom na hierarhijo zanimanj.
- Število poizvedb po posameznih UDK področjih z ozirom na hierarhijo zanimanj.
- Vsota dejavnika K po UDK področjih z ozirom na hierarhijo zanimanj.

Na podlagi ovrednotenja osnovnih miselnih silnic glede na hierarhijo zanimanj sem dobil naslednje izide:

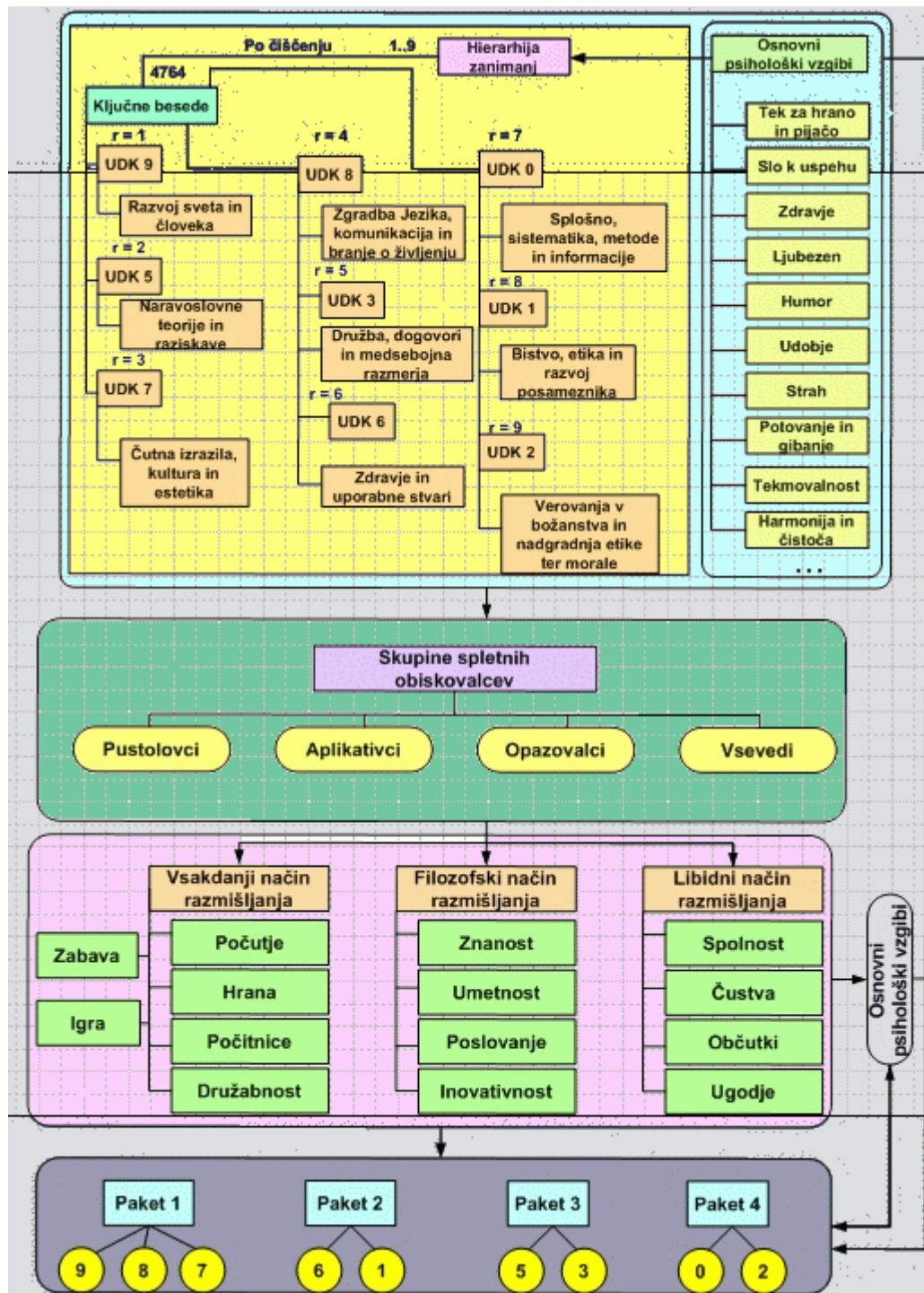
- vodoravno: skupno število točk za vsakdanje razmišljanje je bilo 10, pri čemer sta bila paketa 1 in 2 ovrednotena z oceno 3, medtem ko sta bila paketa 3 in 4 ovrednotena z oceno 2. Skupno število doseženih točk po ovrednotenju za filozofski način razmišljanja je bilo 12 (*vsi paketi so bili ovrednoteni z oceno 3*). Skupno število doseženih točk za libidni način razmišljanja je bilo 6, pri čemer sem paket 1 ovrednotil z oceno 2, paket 2 z oceno 3, paket 3 z oceno 1, medtem ko sem paketu 4 dodelil najnižjo oceno tj. 0.

- Navpično: paket 1 je glede na osnovne miselne silnice dosegel osem točk. Druge značilnosti paketa 1 je možno razbrati s preglednice 8 (*število obiskov je bilo 41394, število poizvedb je bilo 5538 in vsota dejavnika K je bila 3385,44*). Paket 2 je z ozirom na osnovne miselne silnice bil ovrednoten z devetimi točkami (*število obiskov je bilo 15653, število poizvedb je bilo 2763 in vsota dejavnika K je bila 637,77*). Paket 3 je glede na osnovne miselne silnice dosegel šest točk (*število obiskov je bilo 21147, število poizvedb je bilo 3689 in vsota dejavnika K je bila 887,26*). Nenazadnje je bil paket 4 ovrednoten s petimi točkami (*število obiskov je bilo 15616, število poizvedb je bilo 1469 in vsota dejavnika K je bila 480,41*).

Desna stran matrike prikazuje izpeljane oziroma razčlenjene semantične trditve za pakete oziroma za UDK področja, ki hkrati pomenijo tudi temeljni označevalec za štiri skupine spletnih obiskovalcev, medtem ko pomenijo osnovne miselne silnice (*posredno osnovni psihološki vzgibi*) spletnih obiskovalcev/poizvedovalcev dodatno označitev le-teh, ki so v ožji povezavi s semantičnimi trditvami (*semantične trditve oziroma temeljne izpeljane značilnosti spletnih obiskovalcev v svoji prvinski osnovni izhajajo iz osnovnih miselnih silnic in s tem posledično iz osnovnih psiholoških vzgibov le-teh*). Podatki kot so število obiskov, število poizvedb in vsota dejavnika K po posameznih paketih dodatno označujejo velikost in pomembnost teh paketov oziroma skupin spletnih obiskovalcev/poizvedovalcev. S tega izhajajoč je možno trditi, da je skupina spletnih obiskovalcev/poizvedovalcev, za katero velja semantična trditev iz paketa 1 največja skupina, ki je najbolj dejavna z vidika uporabe notranjega iskalnika in katera je uporabila pri poizvedbah najbolj pomembne ključne besede. Druga največja in pomembna skupina spletnih obiskovalcev/poizvedovalcev predstavlja tista skupina, za katero velja semantična trditev iz paketa 3. Tretjo največjo in pomembno skupino spletnih obiskovalcev/poizvedovalcev predstavlja tista skupina, za katero velja semantična trditev iz paketa 2, medtem ko se je skupina spletnih obiskovalcev/poizvedovalcev, za katero velja semantična trditev iz paketa 4 uvrstila kot najmanjša in po ključnih besedah izhajajoč najmanj pomembna množica. Na podlagi izidov je potrebno zdaj skupine spletnih obiskovalcev/poizvedovalcev še ustrezno poimenovati. Skupine spletnih obiskovalcev/poizvedovalcev, za katere veljajo semantične trditve bom poimenoval kot:

- paket 1: »Pustolovci« (*najbolj pogosti in najbolj dejavni poizvedovalci*)
- paket 2: »Aplikativci« (*po pogostosti in dejavnosti na tretjem mestu*)
- paket 3: »Opazovalci« (*po pogostosti in dejavnosti na drugem mestu*)
- paket 4: »Vsevedci« (*po pogostosti in dejavnosti na četrtem mestu*)

Prav s poimenovanjem različnih skupin spletnih obiskovalcev sem tako zaključil postopek segmentacije le-teh. Zaradi nazornejšega poznavanje njihovih miselnih svetov bom njihove značilnosti prikazal še slikovno.



**4.9.9.3 Slikovni prikaz 36: Značilnosti miselnih svetov spletnih obiskovalcev/poizvedovalcev**

Izhodišče slikovnega prikaza 36 je slika 14 (gl. opis na strani 73). Slika 36 prikazuje značilnosti miselnih svetov spletnih obiskovalcev/poizvedovalcev. Z ozirom na to, da so

bile na predhodnih straneh natančno opisani psihološki vzgibi, načini razmišljanja in paketi tega na tem mestu ne bom ponovil. Slika 36 je v bistvu pregledne narave, pri čemer želim izpostaviti skupine spletnih obiskovalcev/poizvedovalcev in njihove lastnosti. Te skupine imajo naslednje lastnosti (*navajam lastnosti po velikostnem vrstnem redu*):

- 1.) Pustolovci – ta skupina je kot že omenjeno največja. Tovrstni spletni obiskovalci/poizvedovalci so na notranjem iskalniku UDK leksikona izvedli največ poizvedb (*so bili najbolj dejavni*) in povrh tega so se njihove uporabljene ključne besede uvrstile med najpomembnejše (*npr. zgodovina, Bach, Haydn, Napoleon idr.*). Z vidika osnovnih miselnih silnic je za te spletne obiskovalce/poizvedovalce značilno, da pri poizvedovanju še zlasti izhajajo iz vsakdanjega in filozofskega načina razmišljanja in redkeje iz libidnega načina razmišljanja. Globalno gledano se zanimajo še zlasti za tematike, ki so v povezavi z zabavo, igro, počutjem, hrano, počitnicami, družabnostjo, umetnostjo, znanostjo, poslovanjem in inovativnostjo. S tega je možno ugotoviti povezavo s psihološkimi vzgibi. Za tovrstne spletne obiskovalce/poizvedovalce so zelo značilni psihološki vzgibi kot npr. radovednost, veselje do lova, estetika/harmonija, humornost, so ambiciozni (*slo k uspehu*), tekmovalnost, veselje do športa in igre. Po drugi strani so nekoliko manj čustveni, lojalni oziroma zvesti in dajejo nekoliko manj poudarka na spolnost, obleko, hrano in pijačo. Z vidika UDK področij so tovrstni spletni obiskovalci/poizvedovalci iskali največkrat informacij s področja UDK 9, UDK 8 in UDK 7 (*gre za specializirane tematike s področja geografije in zgodovine držav, biografij znamenitih osebnosti in družin, pomembnih svetovnih zgodovinskih dogodkov, pomembnih književnih in glasbenih del itd.*).
- 2.) Opazovalci – ta skupina je bila druga največja. Tovrstni spletni obiskovalci/poizvedovalci so bili zelo dejavni, vendar manj kot malo prej omenjeni. Njihove ključne besede so se na lestvici uvrstile nižje kot pa ključne besede, ki so jih uporabili »Pustolovci«, vendar pa višje kot pa ključne besede »Aplikativcev in Vsevedov«. Z vidika osnovnih miselnih silnic je za te spletne obiskovalce/poizvedovalce značilno, da pri poizvedovanju še zlasti izhajajo iz filozofskega načina razmišljanja, redkeje iz vsakdanjega načina razmišljanja in še redkeje iz libidnega načina razmišljanja. Globalno gledano se zanimajo še zlasti za tematike, ki so v povezavi z znanostjo, umetnostjo, s poslovanjem in

inovativnostjo, manj se zanimajo za igre, počutje, prehrano, počitnice, družabnostjo in še manj za tematike v zvezi s spolnostjo, čustvi, ugodjem in občutki. Za tovrstne spletne obiskovalce/poizvedovalce so zelo značilni psihološki vzgibi kot npr. radovednost, vedoželjnost, estetika/harmonija, konstruktivnost, so ambiciozni (*slo k uspehu*), tekmovalnost, lojalnost/zvestoba in čistoča. Po drugi strani so nekoliko manj čustveni in dajejo malo poudarka na spolnost, čustev, obleko, hrano in pijačo. Z vidika UDK področij so tovrstni spletni obiskovalci/poizvedovalci iskali največkrat informacij s področja UDK 5 in UDK 3 (*gre za specializirane tematike s področja matematike, statistike, ekologije, varstvo naravnega in družbenega okolja, fizike, migracije prebivalstva, natalitete in mortalitete prebivalstva itd.*).

- 3.) Aplikativci – po velikosti in pomembnosti se je ta skupina spletnih obiskovalcev/poizvedovalcev uvrstila na tretje mesto. Z vidika osnovnih miselnih silnic je za te spletne obiskovalce/poizvedovalce značilno, da pri poizvedovanju enako močno izhajajo iz filozofskega, vsakdanjega in libidnega načina razmišljanja. Globalno gledano se zanimajo še zlasti za tematike, ki so v povezavi z znanostjo, umetnostjo, s poslovanjem in inovativnostjo, za igre, počutje, prehrano, počitnice, družabnostjo, spolnostjo, čustvi, ugodjem in občutki. Za tovrstne spletne obiskovalce/poizvedovalce so zelo značilni psihološki vzgibi kot npr. radovednost, vedoželjnost, konstruktivnost, ambicioznost (*slo k uspehu*), tekmovalnost, lojalnost/zvestoba in čistoča. Z razliko od prvih dveh skupin tovrstni spletni obiskovalci/poizvedovalci dajejo tudi velik poudarek na spolnost, čustva, obleko, hrano in pijačo. Z vidika UDK področij so tovrstni spletni obiskovalci/poizvedovalci iskali največkrat informacij s področja UDK 6 in UDK 1 (*gre za specializirane tematike s področja medicine dela, proizvodnje, organizacije proizvodnega dela, transporta, elektrotehnike, strojništva, gradbeništva, industrije, filozofije sistemov, filozofije življenja, psihologije dela, stresa itd.*).
- 4.) Vsevedi – ta skupina je z glede na predhodne najbolj šibka in najmanj zastopana. Tovrstni spletni obiskovalci/poizvedovalci so bili z vidika poizvedovanja najmanj dejavni. Z vidika osnovnih miselnih silnic je za te spletne obiskovalce/poizvedovalce značilno, da pri poizvedovanju močno izhajajo iz filozofskega načina razmišljanja, manj iz vsakdanjega in še manj iz libidnega načina načina razmišljanja. Globalno gledano se zanimajo še zlasti za

tematike, ki so v povezavi z/s znanostjo, umetnostjo, poslovanjem in inovativnostjo, manj se zanimajo za igre, počutje, prehrano, počitnice, družabnost, spolnostjo. Za tovrstne spletne obiskovalce/poizvedovalce so zelo značilni psihološki vzgibi kot npr. vedoželjnost, konstruktivnost, ambicioznost (*slo k uspehu*), tekmovalnost, lojalnost/zvestoba in čistoča. Z razliko od prvih treh navedenih skupin tovrstni spletni obiskovalci/poizvedovalci dajejo zelo majhen poudarek na spolnost, čustva, obleko, hrano in pijačo. Z vidika UDK področij so tovrstni spletni obiskovalci/poizvedovalci iskali največkrat informacij s področja UDK 0 in UDK 2 (*tudi ta naveza je v bistvu šibka – gre za specializirane tematike kot so npr. sistematika znanosti, metodologija, opredelitve, pismenstvo, informatika, računalništvo, bibliotekarstvo, muzeji, serijske publikacije, slovarje, leksikoni, kataloge, krščanska in nekrščanska verstva, mitologija, biblijske in mitološke osebe itd.*). Tovrstni spletni obiskovalci/poizvedovalci so največkrat iskali kratke opredlitve s spletnih leksikonov, slovarjev in enciklopedij.

Ob opisih teh štirih skupin spletnih obiskovalcev/poizvedovalcev naj bo ponovno izrazito izpostavljeno dejstvo, da je večina le-teh poizvedovalo na podlagi šolskih (*manj visokošolskih*) programov (*npr. iskanje informacij v zvezi z referatom o Prešernu, Aškercu, Vodniku, Trubarju, s seminarsko nalogo v povezavi s stresom, s seminarskimi nalogami v povezavi z znanimi klasičnimi skladatelji, z domačimi nalogami o književnih in umetniških smereh, z arhitekturo, s seminarskimi nalogami o osmozi, mehaniki, verskimi skupnostmi, onesnaževanju okolja, z informatiko in računalništvom, s filozofijo in filozofov, še zlasti antičnih itd.*). Prav tako je ob tem dejstvu potrebno ponovno izpostaviti drugo nič manj pomembno dejstvo, da je bil spletni UDK leksikon izdelan v namen podpore še zlasti srednješolskega vzgojnoizobraževalnega procesa! To pomeni, da v bistvu v veliki meri ponuja podatke/informacije za srednješolce in srednješolske profesorje, kar s tem posledično sonarekuje njegovo vsebinsko zgradbo. S pomočjo prikazane segmentacije spletnih obiskovalcev/poizvedovalcev sem v bistvu poleg njihovega zanimanja izpeljal njihove osnovne težnje oziroma vzgibe, s katerimi sem lahko dodatno osvetlil njihova psihološka ozadja. To pomeni, da je po moji oceni možno tovrstno segmentacijo spletnih poizvedovalcev uporabiti ne samo v namen podpore vzgojnoizobraževalnega procesa, temveč tudi v smeri trženja tj. za npr. ugotavljanje določenih vedenjskih vzorcev potencialnih potrošnikov določenih proizvodov in storitev.

## 5 Kvalitativni model adaptivne digitalne knjižnice in ovrednotenje

### 5.1 Možnost uporabe izpeljanih novih znanj/spoznanj

Na podlagi s predhodnih poglavij in podpoglavij izpeljanih ugotvitev ter izpeljanih novih znanj/spoznanj v podatkih bom sedaj lahko razvil kvalitativni model adaptivne digitalne knjižnice z dodatnim modulom multidimenzionalnega rang tezavra, ki naj bi se prilagodil še zlasti miselnim hierarhijam spletnih obiskovalcev/poizvedovalcev. O digitalnih knjižnicah in tezavrih sem že poročal v uvodu in drugem poglavju tega doktorskega dela ter ob tem ugotovil, da so tovrstni informacijski sistemi z dodatnimi moduli za upravljanje z znanjem sicer odličen pripomoček za iskanje relevantnih podatkov/informacij, vendar jim po moji oceni primanjkuje prav dve poglobilni značilnosti:

- zmožnost se prilagajati miselnim hierarhijam ljudi (*besedni in grafični vmesnik, s pomočjo katerega bi lahko spletni obiskovalci vrtali po drevesni zgradbi znanja navzgor, navzdol, asociacijsko, ekvivalentno, polarizacijsko in polihierarhično*),
- sposobnost učenja/predvidevanja (*adaptivna digitalna knjižnica, ki bi imela vgrajena dodaten modul za personalizacijo spletnih obiskovalcev in bi na podlagi njihovih uporabljenih ključnih besedah pri poizvedovanju bila sposobna slikovno upodobiti miselne hierarhije za slehernega posameznika in za skupine. Povrhu tega bi se takšen slikovni prikazovalec nenehno posodabljal, tako da bi se s tem tudi drevesna hierarhija povpraševanega človekovega znanja temu ustrezno dopolnila oziroma spremenila*).

Odprava obeh navedenih ključnih pomanjkljivosti adaptivnih digitalnih knjižnic po moji oceni ne more biti več zgolj v domeni bibliotekarske znanosti in stroke, ampak bi se morale najti ustrezne rešitve znotraj aplikativnih ved kot sta še zlasti računalništvo in informatika.

V tej povezavi je potrebno dodati, da bi morala adaptivna digitalna knjižnica imeti sposobnost analize velikih količin podatkov, nakar tudi sintezo dobljenih izidov v oprijemljive podatkovne vzorce, kajti podatki so sami po sebi dokaj neoprijemljive narave.

Ocenjujem, da je potrebno podatkovnim vzorcem, ki so že razvrščeni v posamezne razrede, določiti vizualno obliko in znotraj te oblike morajo biti vzpostavljene ustrezne bolj ali manj dinamične zgradbe ali strukture, kajti oblika sama še ne vsebuje strukture, temveč zgolj določi neko ohlapno mejo do drugih teles oziroma svetov (*podobno kot npr. pri oblikah držav na zemljevidu Evrope, na katerem so države upodobljene v različnih barvah*). Različnim podatkovnim vzorcem z različnih področij človekovega znanja je potrebno na podlagi meritev, opazovanja in/ali izračunov določiti obliko, tako da bi s tem dobili kot izid neko oprijemljivo predstavo. Navedeno je možno prenesti tudi na različne vizualne oblike in strukture poizvedovanj spletnih obiskovalcev oziroma še zlasti njihovih skupin. Skratka shranjevanje in dinamično spreminjanje tovrstnih oblik ter struktur, bi bil eden od temeljnih pogojev inteligentnega učenja.

Adaptivno digitalno knjižnico, kot si jo predstavljam naj bi bila sposobna shranjevati veliko količino podatkov (*takšen modul bi bil primerljiv s spominom*), naj bi bila sposobna spremljati/personalizirati dejavnosti spletnih obiskovalcev (*ta trackerski modul/senzor bi lahko bil pogojno primerljiv s čutili*) in nenazadnje naj bi bila zmožna prepoznati različne vzorce in zgradbe v podatkih oziroma sledovih, ki so jih za sabo pustili spletni obiskovalci, kar bi bil predpogoj za inteligentno učenje in kognicijo (*tovrsten najbolj kompleksen modul, bi bil pogojno primerljiv s kognitivnim aparatom v človekovih možganih*). Iz velike množice pridobljenih podatkov o spletnih obiskovalcih/poizvedovalcih naj bi določen modul adaptivne digitalne knjižnice bil sposoben izdelati nek vizualni svet, ki bi bil pogojno primerljiv z organizmom in znotraj tega organizma naj bi prikazal posebne oblike in strukture, ki so delčki tega velikega organizma (*gl. zamisel na strani od 68 do 69*). S tem posledično bi bilo možno spletne obiskovalce/poizvedovalce tako razvrstiti v ustrezne interesne/značajske skupine kot tudi v nadaljevanju ob njihovih nadaljnjih poizvedovalnih dejavnostih tudi prepoznati (*prepoznavanje in razvrščanje spletnega obiskovalca/poizvedovalca v določeno interesno skupino bi bilo možno na podlagi IP naslova in njegove dosedanje ustvarjene pojmovne mreže, ki je bila izid njegovih poizvedb*).

Ugotavljanje glavne ideje spletnega obiskovalca/poizvedovalca in/ali skupine spletnih obiskovalcev/poizvedovalcev bi dodatno izboljšalo kakovost oziroma relevantnost posredovanih podatkov/informacij, ki bi jih lahko adaptivna digitalna knjižnica z ustreznimi moduli kot izid izbrala in v nadaljnji stopnji povpraševalcem ponudila. »Ena ideja ali zamisel je večinoma nekakšen glavni junak v zgodbi, kjer so vpletene bolj ali manj številne

ideje.«<sup>41</sup> Zgrajena pojmovna mreža določenega poizvedovalca je izid razvoja, ki pa izhaja iz najbolj superiorne- nadrejene zamisli, ki je sestavljena iz najbolj pomembnih pojmov, ki se znotraj celovite pojmovne mreže večkrat pojavljajo (*pogojno primerljivo z glavnim junakom v romanu, ki se npr. na 500 straneh najbolj pogosto pojavlja in ki je v zvezi z izidom ali koncem romana najbolj ključen ali npr. pogojno primerljivo s posameznikom, ki zasleduje v svojem življenju ključno zamisel, da bi si ustvaril družino in šele na naslednjem miselnem hierarhičnem nivoju se nahajajo zamisli o zaslužku, o dobrem avtomobilu idr.*). Naslednji modul, ki bi bil iz mojega vidika pomemben, bi lahko bil predlagatelj za pripravo »pametnih« podatkov, saj bi na podlagi različnih kazalcev (*npr. dolžina spletne seje, uspešnost poizvedb, uporabljene ključne besede, obstoječe pojmovne mreže idr.*) pripravil predlog, katere podatke/informacije bi bilo dobro na določeni zbirki znanja postaviti skupaj.

Nenazadnje bi kot pomemben modul adaptivne digitalne knjižnice izpostavil multidimenzionalni rang tezaver (*v nadaljevanju MRT*), ki naj bi uporabnikom olajšal tako iskanje in organizacijo informacij kot tudi omogočal povezave za kakovostno komunikacijo med njimi. Prav modul MRT bom v nadaljevanju znotraj kvalitativnega modela adaptivne digitalne knjižnice natančneje predstavil. Prav ta modul je tudi osrednjega pomena tega poglavja. Preden izpostavim modul MRT, naj še zaradi večje nazornosti v slikovnem prikazu in kratkem opisu predstavim prvine, ki bi jih lahko adaptivna digitalna knjižnica (*v nadaljevanju ADK*) vsebovala.

---

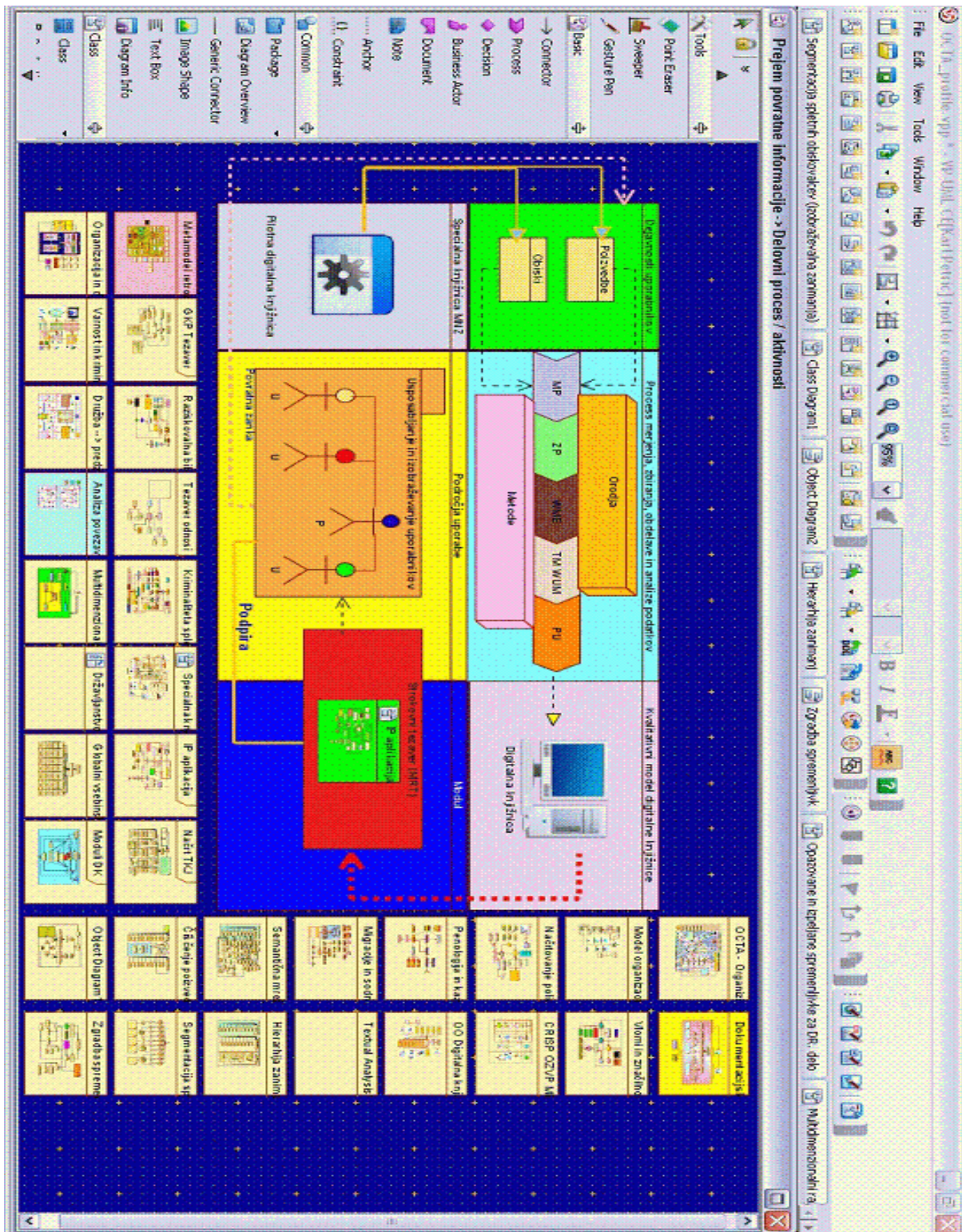
<sup>41</sup> Zamisel sem izpeljal na podlagi svojega eseističnega dela: Petrič, K.(1995). Hierarhologija : 1. del. Celje: Samozal. K. Petrič, 292 str. na strani 10. Delo je nastajalo v letih od 1991 do 1994.



Slika 37 prikazuje možno različico ADK, ki izhaja iz uporabnika in upravljavca elektronskih dokumentov, v katerih so vključeni uporabne informacije za odločitve tako med sodobnimi viri kot tudi v digitalnem (*spletnem*) arhivu. Moduli, ki bi jih lahko sodobna ADK vključevala so po mojem predlogu lahko naslednji:

- diskusijski forumi/komune, v katerih se lahko uporabniki in upravljavci ADK vključijo posamično ali pa znotraj interesnih skupin z namenom, da si izmenjajo izkušnje, znanja in različne individualne zbirke podatkov, s katerimi upravljajo znotraj ADK uporabniki.
- modul OZVP uporabnikov, s katerim se spremljajo dejavnosti in izdelki uporabnikov z namenom, da pridobimo o njih nova znanja in s tem posledično napovedi o pomembnih gibanjih/dogodkov v prihodnosti.
- modul inteligentni bibliotekar, ki svetuje uporabnikom v obliki prijaznega grafičnega vmesnika tako na področju iskanja informacij kot tudi na področju izobraževanja, saj lahko uporabnika napoti na ustrezno diskusijsko komuno, katero obiskujejo tudi ustrezni izvedenci. Inteligentni bibliotekar posreduje uporabnikom avtomatizirane odgovore 24 ur na dan. Povrhu tega je takšen inteligentni bibliotekar tudi zmožen spremljati in shranjevati različne informacijske zahteve uporabnikov, katerim se kot pomemben modul ADK prilagaja.
- modul strukturnik datotek uporabnikom omogoča, da si znotraj ADK vzpostavijo osebni informacijski sistem, s katerimi se lahko povezujejo z drugimi uporabniki in upravljavci ADK. S tem modulom si lahko uporabnik organizira svoje dokumente tako, da vzpostavi med njimi povezave in lahko lažje najde zaželene informacije bodisi znotraj osebnega informacijskega sistema bodisi zunaj le-tega (*brskanje po drugih osebnih informacijskih sistemih s pomočjo dodeljenega gesla ali pa po ADK, ki brez dovoljenja avtorja drugega osebnega informacijskega sistema tovrstne informacije ne vključuje*).
- modul za odločitvena drevesa npr. omogoča najti pravo informacijsko rešitev, da rešimo pereč problem, kajti uporabnik se lahko na podlagi ovrednotenih vej odloča o teži določene rešitve.
- modul inteligentni iskalnik, ki deluje na principu inteligentno pripravljenih podatkov in algoritmov, tako da lahko uporabniki lažje najdejo ne zgolj podatke/informacije, ampak tudi že pripravljene rešitve za določene probleme.
- modul MRT, o katerem bom (*kot že večkrat omenjeno*) v nadaljevanju natančneje poročal v obliki kvalitativnega modela ADK. Kvalitativni model ADK kot tudi izdelan prototip MRT-ja bo tudi dosegljiv na priloženem CD-ROM tega doktorskega dela.

## 5.3 Adaptivna digitalna knjižnica in multidimenzionalni rang tezaver (MRT)



5.3.1 Slikovni prikaz 38: Kvalitativni model adaptivne digitalne knjižnice

Slika 38 v strnjeni obliki ponazarja vse postopke in procese, ki sem jih izvedel, da sem lahko razvil kvalitativni model ADK, ki izpostavlja pomemben modul multidimenzionalnega rang tezavra (*MRT*) za podporo poslovnih in drugih uporabnikov v vzgojnoizobraževalnem procesu. Posamični mejniki (*gl. slika 38*) so oštevilčeni od 1 do 6. Naj na kratko povzamem bistvo tovrstnih mejnikov:

Mejnik 1 – sociološko proučevanje srednješolskega zavoda z vidka bibliotekarske stroke in prakse, izdelava načrta, nakar vzpostavitev pilotne digitalne knjižnice.

Mejnik 2 – spremljanje obiskov in poizvedb na pilotni digitalni knjižnici, nakar klasifikacija poizvedb po UDK področjih.

Mejnik 3 – ponazoritev procesov od merjenja (*MP*), zbiranja (*ZP*), obdelave/priprave (*klasifikacija, WME – webometrične analize idr.*), analize podatkov (*TM WUM – odkrivanje zakonitosti v podatkih in besedilih oziroma poizvedbah*), nakar iz dobljenih novih znanj spoznanj profiliranje uporabnikov (*PU*).

Mejnik 4 in mejnik 5 – kakovosten model ADK, ki vključuje (*uporabniškega*) *MRT* (*ogled aplikativnega prototipa bo možen na priloženem CD-ROM v obliki prikazanega modela s pomočjo uporabe programskega orodja UML Visual Paradigm 6.0*).

Mejnik 6 – prikaže področje uporabe (*uporabniškega*) *MRT* v vzgojnoizobraževalnem procesu. V naslednjem podpoglavju bom natančneje obrazložil zgradbo in cilj *MRT*-ja.

## **5.4 Adaptivna digitalna knjižnica (*ADK*) in multidimenzionalni rang tezaver (*MRT*) in ovrednotenje modela<sup>42</sup>**

DK lahko pomeni odlična platforma za profiliranje uporabnikovih izobraževalnih interesov, kajti na podlagi izdelanih informacijskih/interesnih profilov uporabnikov je možno izdelati DK, ki se prilagaja in učinkovito podpira različne vrste uporabnikov. DK, ki se prilagaja informacijskim potrebam uporabnikov imenujemo ADK. ADK se lahko uporablja v različnih delovnih/poslovnih okoljih (*npr. vladne institucije, znanstvene raziskovalne institucije, vzgojnoizobraževalni zavodi, industrija itd.*). ADK si zamišljam kot aplikativni sistem, kateri zmore kakovostno ter učinkovito podpreti vzgojnoizobraževalne in znanstveno raziskovalne procese ter s tem posledično tudi odgovarjajoče uporabnike.

---

<sup>42</sup> Na podlagi izvedenih poizvedb na različnih specializiranih zbirk podatkov od Web of Science, INSPEC idr. sem ugotovil, da se izraz *MRT* v znanstveni strokovni literaturi izjemno redko pojavlja. Seznam tovrstnih zadetkov bom navedel na koncu tega dela kot priporočene vire.

ADK z dodatnim modulom MRT (*še zlasti njegovih dimenzij*), je prilagojena uporabnikom vzgojnoizobraževalnega procesa v srednjih šolah ter jim na podlagi že pripravljenih informacij olajša iskanje in jim prihrani mnogo časa.

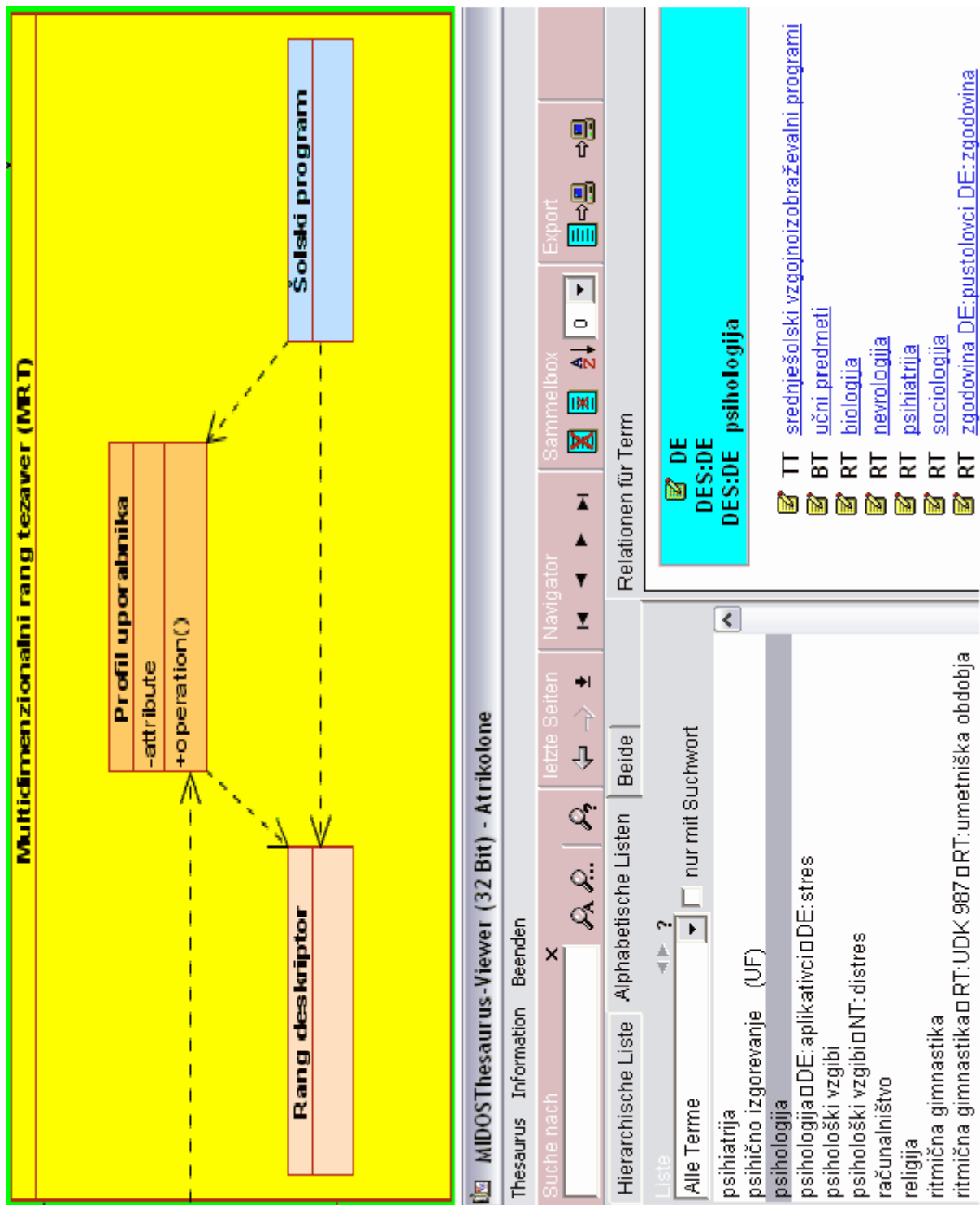
ADK z dodatnim modulom MRT je izid večletnih raziskovanj tako vzgojnoizobraževalnega procesa kot tudi dejavnosti na spletu. Uporabniki lahko pri iskanju ustreznih informacij uporabijo navadni indeksni iskalnik ali pa uporabniški tezaver, v katerem so ključne besede rangirane na podlagi preteklih raziskovanj in na podlagi profilov uporabnikov. Uporabniki se lahko znotraj tezavra vključijo v razne druge storitve, ki jih ponuja pričujoča ADK tudi znotraj MRT kot npr. na različne specializirane diskusijske forume, spletne krožke idr. Storitve so številne kot se prav gotovo za sodobno in kakovostno ADK z MRT spodobi. Te se raztezajo od sodelovanja v različnih specializiranih diskusijskih forumih, spletnih krožkih, raziskovalnih dejavnostih, ustvarjanja lastnih osebnih informacijskih sistemov idr. znotraj digitalne knjižnice. Spremljajo se prav tako zanimanja uporabnikov, tako da se jim digitalna knjižnica nenehno prilagaja.

Uporabniški MRT že vsebuje pomembne ključne besede, po katerih so uporabniki v preteklosti najpogosteje povpraševali in povrh tega so v njem že ustvarjene povezave z drugimi pojmi, ki so vsebinsko bolj ali manj v ožji povezavi. Iskanje znotraj uporabniškega tezavra je zelo enostavno in zelo pregledno. Deluje na principu, da so informacije pripravljene tako, da je čim več koristnih informacij za določenega profila uporabnika na enem mestu. Sicer so informacije tudi razvrščene po sistemu UDK, kar dodatno olajša iskanje informacij, še zlasti bibliotekarjem.

V MRT so informacije organizirane na osnovi interesnih profilov uporabnikov, znotraj katerih so prikazani različni odnosi med deskriptorji (*hierarhija, ekvivalentnost, asociativnost, sinonimnost, antonimnost, homonimnost*).

Obravnavani MRT vključuje tri različne dimenzije, ki so si med sabo tudi v ožji povezavi in so naslednje:

- program srednješolskega vzgojnoizobraževalnega procesa
  - profilirani uporabniki
  - rangirani deskriptorji, ki so izid poizvedb na notranjem iskalniku pilotne digitalne knjižnice
- Naj zaradi večje nazornosti še slikovno prikažem posamezne dimenzije, njihove medsebojne odnose in v končni stopnji še izsek prototipa MRT, ki sem ga izdelal s pomočjo programskega orodja za izdelovanje tezavrov MIDOS 2000 (*prototip tezavra si bo bralec prav tako lahko ogledal na priloženem CD-ROM*).

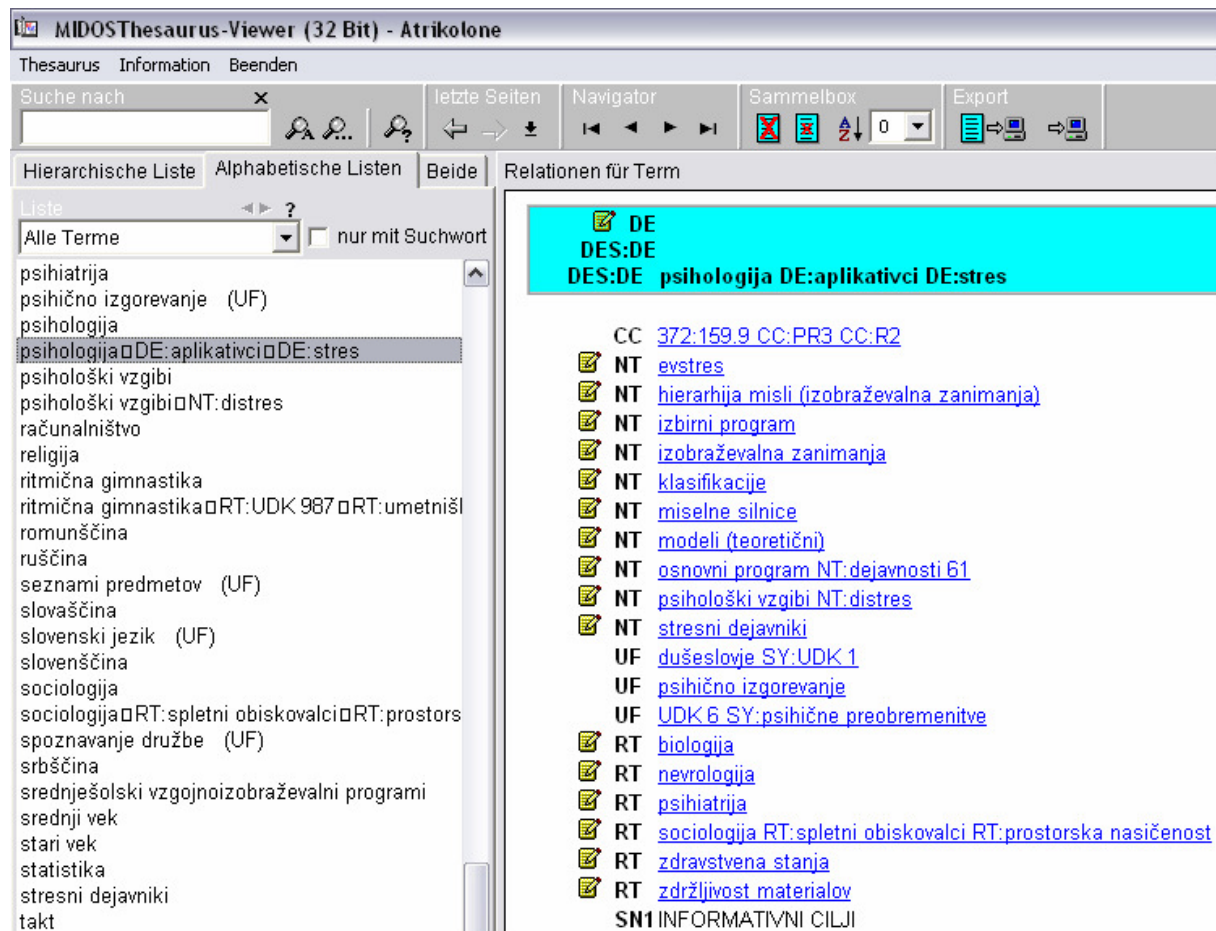


## 5.5 Slikovni prikaz 39: Zgradba multidimensionalnega tezavra (MRT) in izsek prototipa tezavra

Slika 39 prikazuje zgradbo oziroma odnose med dimenzijami MRT in izsek prototipa, ki sem ga izdelal (*kot že omenjeno*) s pomočjo programskega MIDOS 2000. V MRT sem uvrstil dimenzijo ključnih besed in/ali deskriptorjev, ki so na podlagi meritev in analiz dosegle določene range in dejavnike **K**, dimenzijo profiliranih uporabnikov, ki so v proučevanih

obdobjih pri poizvedbah pogostokrat uporabili določene ključne besede in nenazadnje zelo pomembno dimenzijo šolski program, pod katerim vplivom so bile mnoge poizvedbe izvedene. Na levi strani slike 39 lahko opazimo, da sta dimenziji v zelo veliki meri odvisni od dimenzije šolskega programa, in dimenzija rang deskriptor je povrh tega zelo odvisna od dimenzije profiliranega uporabnika, ki so uporabili ključne besede pri poizvedbah. V kolikor so uporabniki poizvedovali pod vplivom šolskega programa, v toliko so bile uporabljene ključne besede tudi posledica te dimenzije. Po moji oceni je bilo tovrstnih usmerjenih ali pod vplivom kolektivnega programa kar 90 % takšnih poizvedb, medtem ko je bilo okoli 8% do 9% poizvedb takšne narave, da so izhajale iz individualnih teženj uporabnika. Skoraj zanemarljiv odstotek poizvedb je bila v znamenju šuma, kar pomeni, da težko ovrednotimo tovrstne poizvedbe kot ali kolektivno ali individualno usmerjene težnje uporabnikov. Za obravnavani MRT bi lahko pogojno trdil, da je to tezaver, ki vsebuje oziroma povezuje tri različne tezavre v eni podobi. S kavzalnega in etialnega vidika bi lahko trdil, da je dimenzija šolski program nadrejena dimenzijama profil uporabnika in rang deskriptorja, a dimenzija profil uporabnika je nadrejena dimenziji rang deskriptorja, čigar videnje pa ne sovпада z globalnega razvojnega vidika. V tem vpogledu je potrebno poudariti, da sem izpeljal dimenzijo profil uporabnika na podlagi meritev in analiz poizvedb, ki so jih izvedli uporabniki na pilotni digitalni knjižnici, ki pa so v večji meri poizvedovali pod vplivom kolektivnega oziroma šolskega programa. Dimenzije so si v bistvu v nekem polihierarhičnem odnosu, v katerem lahko eksistirajo med sabo ali enakovredno (*ekvivalentno*), relacijsko (*sinonimno, asociacijsko*), antonimno (*nasprotni odnos – negativna korelacija*) ali pa hierarhično (*nadrejeno : podrejeno*). Prav ta raznovrstni spekter odnosov med dimenzijami se lahko spreminja zaradi ciljne usmerjenosti tako poslovnih kot tudi drugih uporabnikov. Informacijski poudarek uporabnika v bistvu tudi določa odnose med dimenzijami, kajti reševanje informacijskega problema se lahko v jedru povsem osredotoči zgolj na eno (*npr. dijak želi rešiti informacijski problem za izdelavo raziskovalne naloge ali želi sodelovati v določeni diskusijski komuni*), dve (*npr. pedagoški kadri proučujejo zanimanja dijakov pri učenju, da bi izboljšali kakovost podajanja učne snovi*) ali pa vse tri dimenzije (*npr. upravljavci ADK celovito proučujejo sistem, da bi rešili zadani informacijski problem in s tem izboljšali aplikacijo oziroma aplikacijski sistem*). Desna stran slike 39 prikazuje fragment prototipa MRT, kjer so npr. uporabniku dane možnosti, da si poišče informacije znotraj vseh treh dimenzij, vključujoče različnih odnosov med njimi. Desna stran slike 39 prikazuje okno drevesne zgradbe in okno polihierarhične zgradbe, kjer so izraženi odnosi med dimenzijami. Dani primer kaže, da si je uporabnik izbral področje psihologije (*gl. levo stran okna na desni*

strani slike 39). Desna stran okna s tem posledično prikaže hierarhične in relacijske odnose do pojma psihologije. V primeru, da bi uporabnik na levi strani okna osvetlil polje »psihologija : aplikativci : stres«, bi dobili kot izid naslednjo situacijo, ki jo fragmentarno prikazuje naslednja slika:



## 5.6 Slikovni prikaz 40: Odnosi med dimenzijami

S slike 40 je možno razbrati, da lahko uporabnik znotraj MRT izbira med naslednjimi možnostmi:

- učni predmet psihologija, ki se umešča znotraj predmetnika v srednješolskem izobraževalnem programu (*dimenzija šolski program*).
- profil uporabnika aplikativci, znotraj katerega se lahko uvrsti, če sodeluje v tovrstnih diskusijskih forumih in/ali komunah zaradi učenja in/ali izmenjave izkušenj ter znanj.
- uporabnik se odloči, da bo natančneje proučil področje stresa, ki spada pod psihologijo in je povezan z drugimi znanostmi in vedami (*npr. z medicino, nevrologijo, psihiatrijo, sociologijo, fiziko*).
- v primeru poslovnih uporabnikov (*pedagoški kadri, bibliotekarji, dokumentalisti, informatiki, administratorji idr.*) in morda nekaterih drugih uporabnikov (*npr. dijaki,*

*pripravniki*), se lahko zgodi, da bodo potrebovali kot že omenjeno dve ali pa vse tri dimenzije. Prav s tega vidika je zelo pomemben podatek na vrhu, ki je označen z oznako CC (klasifikacijska kolona – v angl. *Classification Colone*), kajti informacije znotraj MRT-ja so urejene po sistemu UDK in/ali ustreznih dimenzijah. V danem primeru lahko še zlasti poslovnemu uporabniku klasifikacijska oznaka **CC 372:159.9 CC PR3 CC R2** sporoča naslednje:

a.) Polje »psihologija : aplikativci : stres« je razvrščeno kot učni predmet znotraj vzgojnoizobraževalnega procesa z UDK vrstilcem 372, ki je v tem pogledu nadrejeno področju psihologije z vrstilcem 159.9!

b.) Z oznako CC PR3 so označeni uporabniki, ki so profilirani kot »APLIKATIVCI«, ki so se od štirih profiliranih skupin uporabnikov glede pogostosti in intenzivnosti izvedenih dejanj (*poizvedovanje, obiski, dolžina spletnih sej idr.*) na pilotni digitalni knjižnici uvrstili na tretje mesto in nenazadnje še

c.) CC R2 – ta oznaka uporabniku pove, da je ključna beseda stres od vseh drugih uporabljenih in izmerjenih ključnih besed dosegla drugi rang in s tem drugi najvišji dejavnik **K**, kar pomeni, da je ta ključna beseda zelo pomembna tako za profilirane aplikativce kot tudi po vsej verjetnosti za druge profilirane uporabnike (*pustolovci, opazovalci in vsevedi*). Na podlagi CC-ja je možno takoj ugotoviti ne samo kam določena ključna beseda spada, ampak tudi kdo jo najpogosteje uporablja in kako visoko je rangirana, kar omogoča hkrati tudi napovedovanje prihodnjih zanimanj drugih uporabnikov.

Pri vsem tem je pomembno, da so za uporabnike informacije lokacijsko gledano blizu skupaj, kar olajša tako iskanje informacij kot tudi svetovanje uporabnikom znotraj referenčnega procesa s strani bibliotekarja in/ali dokumentalista – informatika. Prav zdaj lahko še za zaključek tega podpoglavja predstavim namen in cilj MRT:

- da uporabnikom omogoča bolj prilagojen in s tem posledično učinkovitejši dostop do uporabnih informacij (*zapisano velja tako za digitalne knjižnice na spletu kot tudi na CD-ROM*).
- da kot izid iskanja oziroma poizvedovanja daje nižjo stopnjo poizvedovalnega šuma
- da lahko MRT uporabimo kot orodje pri razreševanju informacijskih problemov tako v informacijsko dokumentalističnih centrih kot tudi v knjižnicah (to v *sklopu referenčnega procesa bibliotekar/dokumentalist v odnosu z uporabnikom*).
- da omogoča socialno zblíževanje z drugimi uporabniki, še zlasti sorodnimi

- da deluje kot odlična raziskovalna platforma glede raziskovanja uporabnikov in s tem posledično dodatno pripomore k večji prilagodljivosti informacijske ponudbe in storitev tako na področju bibliotekarskih kot tudi na področju pedagoških storitev!

## 5.7 Preglednica 9: Vrednotenje in kritična presoja modela z ozirom na uporabnika

Pomen znakov	Vrsta uporabnikov						
1 - 5 Najmanjša učinkovitost	Bibliotekarji	Dokumentalisti	Informatiki	Pedagoški uslužbenci	Študentje	Dijaki	Upokojenci
6 - 7 Manjša učinkovitost							
8 Srednja učinkovitost							
9 Boljša učinkovitost							
10 Najboljša učinkovitost							
Usmeritve	Motiviranje uporabnikov						
Kakovost storitev	10	10	10	10	9	9	7
Prirast članstva in boljše sodelovanje	10	10	9	10	10	9	5
Razumljivost	10	10	4	3	3	2	1
Ustvarjanje družbene blaginje	10	10	10	10	10	9	7
Vedenjski vzorci ljudi (proučevanje)	10	10	10	10	10	10	8
Izobraževanje uporabnikov	10	10	10	10	10	9	5
Raziskovanje	10	10	10	10	10	10	7
Kreativnost glede izmenjave informacij	10	10	10	10	10	10	7
Sociabilnost ljudi	10	10	10	10	10	10	8
Uspešnost dela	10	10	8	10	9	9	6
<b>Skupaj</b>	<b>100</b>	<b>100</b>	<b>91</b>	<b>93</b>	<b>91</b>	<b>87</b>	<b>61</b>

Preglednica 9 predstavlja ovrednotenje kvalitativnega modela ADK z modulom MRT na podlagi učinkovitosti usmeritev z ozirom na moč motivacije različnih vrst uporabnikov (*uporabniki so naslednji: bibliotekarji, dokumentalisti, informatiki, pedagoški uslužbenci, študentje, dijaki in upokojenci*). V skladu s tem sem ovrednotil najbolj pomembne usmeritve s pomočjo naslednjih uteži: od 1 do 5 (*pomeni najmanjša učinkovitost*), od 6 do 7 (*pomeni manjša učinkovitost*), utež 8 (*pomeni srednja učinkovitost*), utež 9 (*pomeni boljša*

*učinkovitost*) in nenazadnje utež 10 (*pomeni najboljša možna učinkovitost*). Usmeritve so naslednje:

- kakovost storitev ( *vključuje obstoječe in nove koristne storitve za različne informacijske potrebe različnih vrst uporabnikov*)
- prirast članstva in boljše sodelovanje
- razumljivost ( *vključuje navigacijo in učinkovitost iskanja informacij*)
- ustvarjanje družbene blaginje ( *koristna znanja se v kasnejšem obdobju lahko še nadalje plemenitijo, pri čemer ima družba tudi korist ne samo v smeri duhovnega, temveč tudi v smeri denarnega dobička*)
- proučevanje vedenjskih vzorcev uporabnikov ( *vključuje moč oziroma dozdevno učinkovitost profiliranja uporabnikov*)
- izobraževanje uporabnikov ( *vključuje učinkovitost izobraževalne podpore uporabnikov tudi s pomočjo različnih didaktičnih pripomočkov, modelov ipd.*)
- raziskovanje ( *vključuje učinkovitost spodbujanja raziskovalnega duha pri uporabnikih*)
- kreativnost glede izmenjave informacij ( *upošteva učinkovitost glede na dejavno izmenjavo koristnih informacij med uporabniki*)
- sociabilnost ljudi ( *poudarek je na dozdevni učinkovitosti socialnega zblíževanja uporabnikov s pomočjo socio- programske opreme*)
- uspešnost dela ( *ta usmeritev predvideva večjo učinkovitost dela pri vseh uporabnikih npr. pri pedagoških uslužbencih je to nazornejša predstavitev učne snovi, medtem ko je to pri študentih ali dijakih boljši študijski oziroma učni uspeh ipd.*)

Glede učinkovanja usmeritev na različne vrste uporabnikov sem po ovrednotenju dobil izide v vertikali ( *izidi – navpični pogled oziroma po stolpcih*) in v horizontali ( *izidi – vodoravni pogled oziroma po vrsticah*)

**Ovrednotenje po stolpcih** ( *učinkovitost posameznih usmeritev glede na posamezne vrste uporabnikov*) je dajalo naslednje izide:

- 100 točk od 100 možnih za bibliotekarje in dokumentaliste, kar pomeni, da bi lahko model ADK z MRT v najboljši možni meri spodbudil učinkovitost navedenih usmeritev za delo obeh vrst uporabnikov
- 93 točk od 100 možnih za pedagoške uslužbence pomeni zelo dober izid
- 91 doseženih točk od 100 možnih za informatike in študente
- 87 doseženih točk od 100 možnih za dijake
- najmanjše število točk so dosegli upokoenci tj. 61 točk od 100 možnih

**Ovrednotenje po vrsticah** ( *učinkovitost določene usmeritve z ozirom na vse uporabnike*)

- kakovost storitev: na podlagi izboljšane kakovosti storitve za uporabnike je ta usmeritev dosegla 65 točk od 70 možnih, kar je zelo dober izid
- prirast članstva in boljše sodelovanje: ta usmeritev je dosegla 63 točk od 70 možnih, kar pomeni prav tako zelo dober izid
- razumljivost: ta usmeritev je dosegla zgolj 33 točk od sto možnih, kar je dokaj slab izid (*uporabniki kot so informatiki, študentje, dijaki in upokojeanci bi lahko doživeli ta model ADK z MRT kot manj razumljiv, kar bi lahko pomenilo, da je doba učenja daljša*)
- ustvarjanje družbene blaginje: ta usmeritev je dosegla 66 točk od 70 možnih (*zelo dober izid*)
- proučevanje vedenjskih vzrocev ljudi oziroma uporabnikov: ta usmeritev je dosegla kar 68 točk od 70 možnih (*ta odličen izid bi lahko pomenil možno učinkovito profiliranje uporabnikov, kar je osnovni pogoj za ustrezno delujočo adaptivno digitalno knjižnico*)
- izobraževanje uporabnikov: usmeritev je dosegla 64 točk od 70 točk (*zelo dober izid*)
- raziskovanje ali spodbujanje raziskovalnega duha: 67 točk od 70 možnih (*odličen izid*)
- spodbujanje glede učinkovitosti kreativnosti izmenjave koristnih informacij: odličen izid – od 70 možnih je ta usmeritev dosegla 67 točk
- sociabilnost ljudi ali povezovanje uporabnikov med sabo v koristne interesne in humane skupine: od 70 možnih točk je ta usmeritev dosegla 68 točk, kar je odličen izid
- uspešnost dela oziroma dvig delovne uspešnosti: 62 točk od 70 možnih (*zelo dober izid*)

### **Kritična presoja**

Na podlagi ovrednotenih usmeritev z ozirom na različne vrste uporabnikov sem ugotovil, da je dosegel kvalitativni model ADK z MRT zelo dobre izide. Največjo pomanjkljivost pa je možno zaznati prav pri usmeritvi »razumljivosti« (*ovrednotenje po vrsticah*), kajti če je določena aplikacija manj razumljiva za ključne uporabnike kot so pedagogi, študentje in dijaki, je tovrstna pomanjkljivost dokaj kritična. Glavni razlog za dozdevno nižjo stopnjo razumljivosti modela vidim na področju tezavra. Problem bi bilo možno rešiti tako, da bi izobraževalni zavodi uvedli nov študijski/učni predmet za učinkovito organizacijo in iskanje informacij (*usvajanje iskalnih tehnik, poznavanje koristnih zbirk podatkov, izdelava diagramov, poznavanje modelirnih orodij, izdelava digitalnega tezavra itd.*). Prav s tem obeležjem, bi bila še ta pomanjkljivost odpravljena in povrhu tega bi tako pedagoški uslužbenci kot tudi študentje in dijaki pridobili koristna znanja, ki bodo še zlasti v prihodnosti (*morda*) postala nepogrešljiva!

## 6 Sklepno poglavje

V uvodu sem predstavil problematiko digitalnih knjižnic in zgradbo doktorskega dela, nakar sem v nadaljevanju poročal o cilju, hipotezah in raziskovalnih vprašanjih. V tem poglavju sem si zastavil dve raziskovalni predpostavki in osem raziskovalnih vprašanj. V moji raziskavi sem želel potrditi ali pa ovreči naslednji raziskovalni predpostavki:

- 1.) Informacijski/poizvedovalni vedenjski vzorec spletnih obiskovalcev na obstoječi pilotni digitalni knjižnici, se dolgoročno gledano ne spreminja in je tako neodvisen od različnih časovnih obdobj.
- 2.) Informacijski/poizvedovalni vedenjski vzorec spletnih obiskovalcev na obstoječi pilotni digitalni knjižnici je zelo odvisen od različnih časovnih obdobj in se tako nenehno spreminja.

V poglavju teoretičnih izhodišč sem predstavil področje webometrije (*spletometrija*), ontologije (*znotraj ontologije, še zlasti univerzalni decimalni klasifikacijski sistem in tezaver*) in način razvrščanja poizvedb, ki so jih spletni obiskovalci izvedli (*od 01.01.2003 do 01.01.2006*) na notranjem iskalniku lastne spletne izvedbe UDK leksikona, določene pojme in metode oziroma tehnike s področja (*spletnega*) odkrivanja zakonitostih v podatkih, nakar sem še poročal o digitalnih knjižnicah.

V tretjem poglavju sem predstavil snovanje kvalitativnega modela ADK, nakar sem spletometrične podatke najprej načrtoval (*vkjučuje zbiranje, štetje in razvrščanje podatkov in kot končni izid dobil delno strukturirane podatke*), pripravil (*vključuje integracijo in čiščenje podatkov, določevanje zakona moči, obdelava podatkov s pomočjo programskega orodja Antconc - dobim konstante za UDK področja, pomembne ključne besede, range, frekvence, dejavnike K, nove attribute, skratka kot izid dobim strukturirane podatke*), OZVP (*vključuje besedno analizo, filtriranje besed, rangiranje, štetje, vizualizacijo strukturiranih podatkov s pomočjo metode tj. asociacije in hierarhičnih grozdov idr.*) in ovrednotil s spletometrično metodo zakona moči (*tj. izpeljana različica Zipfovega zakona*). Na podlagi prikazanih spletometričnih meritev, analiz in uporabe določenih metod s področja OZVP sem se lahko dokopal do ugotovitev, s katerimi sem lahko odgovoril na zastavljeni hipotezi in raziskovalna vprašanja, nakar sem izpeljal spoznanja in si s tem posledično pridobil nova zanimiva znanja o zbranih in obdelanih podatkih, ki jih poprej nisem imel. Ob merjenju dejavnosti spletnih obiskovalcev/poizvedovalcev in nadaljnji stopnji obdelavi zbranih podatkov sem imel o

podatkih dokaj omejeno predstavo, kajti ključna in zanimiva dejstva so zame ostala prikrita.

## 6.1 Kakšni so bili izidi?

### 6.1.1 Zakon moči pred in po čiščenju podatkov

S pomočjo metode zakona moči pred in po čiščenju podatkov sem dobil naslednje izide, ki sem jih razdelil v tri velikostne skupine (*podatki od leta 2003 do 2006*):

**Poizvedbe (*rang; UDK; log NP*): Pomembnejše ključne besede (*rang; UDK; log NB*)**

#### 1. ali vrhnja Skupina

Rang 1; UDK 9; 3,4951 : Rang 1; UDK 9; 3,3396

Rang 2; UDK 6; 3,2826 : Rang 2; UDK 5; 2,6910

Rang 3; UDK 5; 3,2702 : Rang 3; UDK 7; 2,5899

#### 2. ali srednja skupina

Rang 4; UDK 3; 3,2619 : Rang 4; UDK 8; 2,5728

Rang 5; UDK 8; 3,0989 : Rang 5; UDK 3; 2,5453

Rang 6; UDK 7; 3,0625 : Rang 6; UDK 6; 2,4668

#### 3. ali spodnja skupina

Rang 7; UDK 1; 2,9283 : Rang 7; UDK 0; 2,4533

Rang 8; UDK 0; 2,9079 : Rang 8; UDK 1; 2,3117

Rang 9; UDK 2; 2,8195 : Rang 9; UDK 2; 2,2810

- 3.) Z ozirom na izide poizvedb in ključnih besed (*gl. levo stran in desno stran*) sem ugotovil, da se »**Informacijski/poizvedovalni vedenjski vzorec spletnih obiskovalcev na obstoječi pilotni digitalni knjižnici, dolgoročno gledano ne spreminja in je tako neodvisen od različnih časovnih obdobj, s čimer sem potrdil prvo zapisano predpostavko, medtem ko sem drugo hipotezo s tem ovrgel.**«

### 6.1.2 Distribucija in statistični atributi za poizvedbe

S pomočjo te vizualizacijske tehnike za prikaz podatkovnih vzorcev za poizvedbe sem ugotovil, da lahko z metodo zakona moči ponovno dokažem stabilnost preučevanega sistema oziroma poizvedovalnega vedenjskega vzorca spletnih obiskovalcev, kajti rangi posameznih UDK področij se niso spremenili. Ko primerjam med sabo frekvence znotraj UDK področij z vsemi frekvencami, dobim glede doseženih rangov posameznih UDK področij isti izid.

### 6.1.3 Primerjava frekvenc znotraj UDK področij z vsemi frekvencami

S slikovno tehniko distribucije sem se lahko dokopal do ugotovitev, ki podkrepijo predhodne izide:

- prevladuje red in stopnja predvidljivosti je večja
- pojav določenih relevantnih ponavljajočih se (*vedenjskih*) vzorcev je pogost
- se ne odzove na majhne spremembe in se zaradi le-teh ne zruši
- vpliv časa je manjši ali celo zanemarljiv
- obstaja nek močan poudarek oziroma izraža neko trdno usmerjenost
- poizvedovalni šum je zanemarljiv

S kakovostnega vidika sem ugotovil naslednje (*uporabil slikovno tehniko distribucije*):

Prihodnostne napovedi krivulj verjetnosti so sorazmerno zvezne, kot da bi imel opraviti z linearnimi trendi, medtem ko kažejo krivulje zanesljivosti ali zaupanja te napovedi nekoliko manj zvezen izraz. Pri tovrstnem kakovostnem ovrednotenju intervalov nisem vključil še količinske kazalce kot so frekvenčni intervali, dejavniki intervalov verjetnosti in zaupanja na merski skali od 0 do 1.

### 6.1.4 Razpon števila poizvedb znotraj UDK področij

Razpon števila poizvedb sem ponazoril s pomočjo slikovne tehnike Scatterplot in s tem dobil naslednje izide:

Spletni obiskovalci so najbolj pogosto izvedli eno določeno poizvedbo v 4865 primerih po enkrat, v 2320 primerih po dvakrat, v 633 primerih po trikrat, v 248 primerih po štirikrat, v 84 primerih po petkrat, v 50 primerih po šestkrat, v 20 primerih po sedemkrat, v 14 primerih po osemkrat, v šestih primerih po devetkrat, v osmih primerih po 10 krat, v dveh primerih po 11 in 12 krat in v enem primeru po 13, 14, 15 in 22 krat. Na podlagi tovrstne porazdeljenosti, ki sem jo prikazal slikovno s črtnim diagramom, sem dobil kot izid skoraj pravilno hiperbolično odvisnost (*podobno kot pri frekvenčni distribuciji besed, gl. Zipfov zakon v teoretičnem uvodu*). Prav podobne izide dobimo, če te porazdelitve prikažemo znotraj posameznih UDK področij. Določena poizvedba, ki je bila večkrat zaporedno izvedena in se v poznejšem obdobju več ne pojavlja, v bistvu ni toliko relevantna kot takšna poizvedba, ki se pojavlja po enkrat ali dvakrat, vendar se pojavlja v različnih časovnih intervalih (*npr. od leta 2003 do 2006*). Dokaz za tovrstno trditev je možno najti še zlasti pri poizvedbah s področja UDK 2.

### **6.1.5 Prikaz podakovnih vzorcev s pomočjo Sievovega diagrama**

S pomočjo slikovne tehnike Sievovega diagrama sem prikazal površine, ki jo zavzemajo poizvedbe po posameznih UDK področjih in pri tem ugotovil naslednje:

- največjo površino zavzemajo poizvedbe s področja UDK 9
- sledita mu področji UDK 6 in UDK 5
- potem UDK 3, UDK 8, UDK 7, UDK 1, UDK 0 in UDK 2.
- zavzeta površina poizvedb s področja UDK 4 je bila na sliki upodobljena kot komajda viden moder kvadrat na poziciji med 6 in 0. Na podlagi ugotovitev sem ustvaril osnovni model sobe ali poslopja, ki sem ga pozneje še naprej razvil.

### **6.2 Odkrivanje zanimivih vzorcev v podatkih z ozirom na poizvedbe s pomočjo dejavnika pomembnosti ali moči določene ključne besede**

Prikazal sem predvsem tiste slikovne prikaze, pri katerih sem primerjal med sabo range in dejavnike **K**.

#### **6.2.1 Distribucija **K**, ranga, statistični atributi dejavnika **K** in stolpčni diagram UDK**

Pri vizualizaciji distribucije **K**, ranga, statistične attribute dejavnika **K** in stolpčnega diagrama za UDK področja sem dobil naslednje izide:

- statistični atributi za dejavnik **K**: vrednosti **K** se gibljejo od 2,09 do 68,65. Povprečna vrednost je od 4,67 +/- 4,75 in od 1155 vrednosti je 80 različnih.
- stolpčni diagram: rangirna lestvica se potem ob upoštevanju pogostosti pojavljanja ključne besede, ki spada v neko določeno UDK področje precej spremenila. Prav to spremembo nam nakazuje upoštevanje ranga (**r**) določene ključne besede.
- distribucija: prikaz dejavnika **K** in **r** za posamezna UDK področja mi je posredovalo zelo zanimivo ugotovitev glede zamenjave ranga za področji UDK 7 in UDK 6, iz česar sem lahko sklepal, da so bile ključne besede s področja UDK 7 višje rangirane kot ključne besede s področja UDK 6, čeprav je bilo več poizvedb s področja UDK 6.

#### **6.2.2 Primerjava dejavnika **K** in **r** s pomočjo Scatterplota**

Ob primerjavi dejavnika **K** (*Y-os*) in **r** (*X-os*) s pomočjo vizualizacijske tehnike Scatterplot (*raztreseni diagram*) sem prispel do naslednjih ugotovitev:

- sorazmerno maloštevilni predstavniki ključnih besed z višjimi vrednostmi dejavnika **K** prispevajo več k skupni vrednosti kot pa sorazmerno množični predstavniki ključnih besed z nižjimi vrednostmi dejavnika **K**.
- v primeru te zastavljene raziskave o poizvedbah na notranjem iskalniku UDK leksikona, je možno poročati o besedah, ki so se pojavile precej pogosto (*npr. v, kako, in itd.*), vendar pa njihov pomen ni bil relevanten. Po drugi strani so se pojavljale določene poizvedbe, ki so bile pogoste in so vsebovale močne ključne besede (*npr. zgodovina, stres, Bach itd.*)
- nekatere druge poizvedbe so bile precej redke, z vidika drugih kontekstov gledano izolirane, tako da so bile te besede po izvedeni besedni analizi ali izločene (*npr. poizvedbe, ki so se pojavile zgolj enkrat ali pa poizvedbe s področja UDK 4*) ali pa so dosegle nizke vrednosti dejavnika **K**.
- kadar gre za življenjepise znamenitih osebnosti z različnih področij znanosti, umetnosti, športa in različnih dejavnosti, je možno sklepati, da v 95 % primerov obstaja močna povezanost do drugega področja človekovega znanja.

### 6.2.3 Vizualizacija ključnih besed na podlagi dejavnika **K**

Na podlagi razvrščenih podatkov s pomočjo dejavnika **K** sem dobil pet različnih velikostnih razredov, ki so naslednji:

1. razred **K** od 68,648 do 68,648: v ta velikostni razred se je uvrstila zgolj ena ključna beseda tj. ZGODOVINA s frekvenco 91 in spada v področje UDK 9.
2. razred od 42,975 do 54,236: v ta velikostni razred se je uvrstila ključna beseda STRES z vrednostjo **K**=54,236 in s frekvenco 53 ter spada v področje UDK 1. Druga ključna beseda, ki se je uvrstila v ta velikostni razred je BACH z vrednostjo **K**=42,975, s frekvenco 31 in spada v področje UDK 9.
3. v razredu s **K** od 31,885 do 41,589 so se uvrstile naslednje ključne besede: UDK z vrednostjo **K**=41,589 (*frekvenca 30, UDK 0*), VIVALDI s **K**=36,044 (*frekvenca 26, UDK 9*), HAYDN s **K** 34,657 (*frekvenca 25, UDK 9*), NAPOLEON s **K**=33,271 (*frekvenca 24, UDK 9*), BEETHOVEN s **K**=33,271 (*frekvenca 24, UDK 9*) in SHAKESPEARE s **K**=31,885 (*frekvenca 23, UDK 9*).
4. v razredu s **K** od 15,904 do 27,726 so se uvrstile ključne besede (*zaradi obširnosti podatkov bom navedel zgolj ključne besede in UDK brez pripadajočih vrednosti*): SELIŠKAR, ANTON, TONE (*vse tri ključne besede spadajo v UDK 9*), SHIZOFRENIJA (*UDK 1*), SCHUBERT, BONAPARTE, GRŠKA, LASSO,

JOSEPH, FRANZ, EGIPT, ANTONIO, WILLIAM, SLOVENIJI, ORLANDO, DALI, CHOPIN (vse UDK 9), SEMINARSKA, SEMINARSKE (UDK 0), IZVOR, RICHARD, PREŠEREN, GRČIJA (UDK 9), FOTOSINTEZA (UDK 5) in nenazadnje GRŠKI (UDK 9).

5. v razredu s **K** od 2,093 do 15,249 se je uvrstilo največje število ključnih besed z različnih UDK področij. Najpomembnejši predstavniki ključnih besed v tem velikostnem razredu so bili naslednji: ZDA (UDK 9), Sebastian (UDK 9), barok (UDK 7), satanizem, hinduizem (UDK 2), klasicizem (UDK 7), ekologija (UDK 5), glasba (UDK 7), Bog (UDK 2), kavčuk (UDK 6), renesanca (UDK 7), Afrika (UDK 9), Islam (UDK 2), angleščina (UDK 8), jehove (UDK 2), Jazz (UDK 7) itd. V tem velikostnem razredu zopet krepko prevladuje področje UDK 9. V tem velikostnem razredu so nekatere ključne besede, ki sem jih razvrstil v področje UDK 7 dosegle sorazmerno visoko vrednost dejavnika **K**, medtem so s področja UDK 6 dosegle nekoliko nižje vrednosti.

#### 6.2.4 Izpeljana asociacijska pravila

Pri minimalni vrednosti 5 % podpore in 50 % zaupanja sem dobil kot izid 80 asociacijskih pravil. V nadaljevanju sem kot minimalno vrednost dejavnika podpore določil 10 %, medtem ko sem za najmanjšo vrednost dejavnika zaupanja določil 100 %. Kot izid sem dobil sedem asociacijskih pravil, ki so naslednji:

biografija = Y glasba = Y -> življenjepis = Y (11.429% 100.00% 4 4) UDK 9

biografija = Y -> življenjepis = Y (28.571% 100.00% 10 10) UDK 9

dela = Y življenjepis = Y -> biografija = Y (17.143% 100.00% 6 6) UDK 9

dela = Y biografija = Y -> življenjepis = Y (17.143% 100.00% 6 6) UDK 9

dela = Y -> življenjepis = Y biografija = Y (17.143% 100.00% 6 6) UDK 9

dela = Y -> biografija = Y (17.143% 100.00% 6 6) UDK 9

dela = Y -> življenjepis = Y (17.143% 100.00% 6 6) UDK 9

S tega izida lahko sklepam, da so najmočnejša asociacijska pravila s področja UDK 9 ali s področja razvoja človeka in Zemlje, natančneje s področja življenjepisov zgodovinskih osebnosti, ki so bili zelo dejavni in za sabo pustili pomembne sledove ter tako bolj ali manj pomembno vršili vpliv na nadaljnji razvoj človeka in posredno tudi Zemlje. V primeru, da ohranimo merila minimalne podpore tj. pri 5 % in zaupanja tj. pri 50 % dobimo zanimiv vpogled v še druge asociacijske povezave, ki povrh tega prikazujejo bolj ali manj močne povezave med različnimi UDK področij, še zlasti so poudarjene povezave

med področij UDK 9, UDK 8 in UDK-jem 7. Ob nadaljnjem poglobljanjem s pomočjo programskega orodja Ilog Discovery sem ugotovil naslednje (*odstotkovni deleži po UDK razvrščenih asociacijskih z ozirom na skupno število asociacijskih pravil, ki je bilo 80*):

- skupni izid je naslednji UDK 9 =  $25 \times 100/80 = 31,25 \%$ ; UDK 8 =  $24 \times 100/80 = 30,00 \%$ ; UDK 7 =  $14 \times 100/80 = 17,50 \%$ ; UDK 6 =  $6 \times 100/80 = 7,50 \%$ ; UDK 5 =  $4 \times 100/80 = 5,00 \%$  in UDK 3 =  $7 \times 100/80 = 8,75 \%$ .

- asociacijskih pravil ni bilo s področja UDK 0, UDK 1 in UDK 2. Še zlasti nekoliko preseneča ugotovitev, da ni bilo s področja UDK 1 nobenih asociacijskih pravil, čeprav je bilo možno na prvih 34 mestih beležiti dve močni ključni besedi kot sta »Stres« in »Shizofrenija«, ki sta se pogostokrat kot poizvedbi pojavili kot enobesedna zveza, kar spada pod UDK 1, vendar v vsebinski relaciji sta se pojavili v povezavi UDK 3 in še zlasti s področjem UDK 6. Ključna beseda »Stres« se je pojavila še zlasti v zvezi s službo, transportom, delovnim mestom in zdravstvom, medtem ko se je ključna beseda »Shizofrenija« pojavila zgolj kot enobesedna zveza in tako zgolj posredno odpirala povezavo na asociacijo psihologija, bolezen in zdravstvo.

S pomočjo slikovne tehnike dveh distančnih map tj. za dejavnika podpore in zaupanja (*merilo: evklidska razdalja*) ter dveh distančnih map za prikaz ocenjene moči asociacijskih povezav in pravil (*merilo evklidske razdalje*) sem ugotovil naslednje:

- najmanjša evklidska razdalja med dvema vrednostima je bila ob pogoju ( $S1 = 96,0 \%$ ,  $C1 = 98,9 \%$ ;  $S2 = 96,0 \%$ ,  $C2 = 98,4 \%$ ), ki je bila 0,0121850725. Gre za že znani izpeljani asociacijski pravili izpeljani iz številčnih podatkov po izvedeni besedni analizi poizvedb  $D\_K \leq 15.404 \rightarrow D\_N \leq 19.80$  (96,0 %, 98,9 %) in  $D\_N \leq 19.80 \rightarrow D\_K \leq 15.404$  (96,0 %, 98,4).

- Evklidske razdalje ovrednotenih asociacijskih povezav med posameznimi UDK področji so najmanjše med UDK 6 in UDK 5 ter UDK 8 in UDK 7 tj. 1,414. Na drugem mestu glede najmanjših evklidskih razdalj so se uvrstile ovrednotene asociacijske povezave med UDK 9 in UDK 8 ter med UDK 9 in UDK 7 (*evklidska razdalja je enaka 2,4494...*).

- pri oceni moči povezave po UDK razvrščenih asociacijskih pravil sem ugotovil, da dobimo najmanjšo evklidsko razdaljo za asociacijska pravila med UDK 9 in UDK 8, med UDK 9 in UDK 7 (*vrednost je 1,414...*). Vrednost evklidske razdalje je enaka 2,0 dobimo za asociacijska pravila med UDK 6 in UDK 5. Moč povezave med asociacijskimi pravili s področja UDK 8 in UDK 7 je enaka 1,732. Nenazadnje obstaja še ena ovrednotena moč relacije za asociacijska pravila med UDK 6 in UDK 3 (*vrednost je enaka 2,0*). **Glavno**

**središče moči povezanosti med asociacijskimi pravili po razvrščenih UDK področjih se nahaja pri UDK 7, 8 in 9.**

### **6.2.5 Hierarhično grozdenje pomembnih ključnih besed**

Vsota dejavnika pomembnih ključnih besed ( $\sum K$ ), odstotkovni deleži pomembnih ključnih besed ( $\%K$ ), vsota števila obiskov ( $\sum No$ ) in odstotkovni deleži števila obiskov ( $\%No$ ) po posameznih UDK področjih so bili naslednji:

$\sum K$  za področje UDK 0 je znašala 272,25 ali 5,05 %,  $\sum No$  je bila 5820, kar odgovarja 6,20 %. Pri področju UDK 1 je bila  $\sum K$  264,48 ali 4,91 %,  $\sum No$  je bila 7479 ali 7,97 %. Pri področju UDK 2 je bila  $\sum K$  208,16 ali 3,86 %,  $\sum No$  je bila 9796 ali 10,44 %.  $\sum K$  za področje UDK 3 je znašala 387,73 ali 7,19 %,  $\sum No$  je bila 10910, kar odgovarja 11,63 %. Pri področju UDK 5 je bila  $\sum K$  499,53 ali 9,27 %,  $\sum No$  je bila 10237 ali 10,91 %.  $\sum K$  za področje UDK 6 je znašala 373,29 ali 6,42 %,  $\sum No$  je bila 8174, kar odgovarja 8,71 %. Pri področju UDK 7 je bila  $\sum K$  318,70 ali 5,91 %,  $\sum No$  je bila 11991 ali 12,78 %. Pri področju UDK 8 je bila  $\sum K$  389,14 ali 7,22 %,  $\sum No$  je bila 13830 ali 14,74 % in nenazadnje za področje UDK 9 je bila  $\sum K$  2677,60 ali 49,67 %,  $\sum No$  je bila 15573 ali 16,60. S pomočjo metode hierarhičnega grozdenja in slikovne tehnike dendrograma sem kot izid dobil naslednje grozde:

- grozd UDK 9 zajema vseh devet točk tj. 9, 8, 7, 6, 5, 4, 3, 2, 1 in 0. Višina grozda je enaka 1,137438.
- grozd UDK 8 zajema pet točk tj. 8, 7, 5, 3 in 2. Višina grozda znaša 0,320138. Grozd UDK 8. Točki 3 in 5 sta med sabo povezani in imata drugo najmanjšo razdaljo, ki znaša 0,082792.
- grozd UDK 0 zajema tri točke tj. 0, 1 in 6. Višina grozda je enaka 0,206706. Točki 1 in 6 sta med sabo povezani in imata najmanjšo razdaljo, ki je enaka 0,078418.

Z vidika velikosti grozda UDK 9 bi lahko trdil, da mu je najbolj soroden grozd UDK 8, ki vključuje UDK 7, UDK 5, UDK 3 in UDK 2, nakar šele sledi grozd UDK 0, ki vključuje UDK 1 in UDK 6. Skratka ob primerjavi dveh izbranih kazalcev tj.  $\%K$  in  $\%No$  (primerjava med odstotkovnim deležem najmočnejših ključnih besed po UDK področjih in odstotkovnim deležem pogostosti obiskov po UDK področjih) sem dobil kot izid nekakšno hierarhijo (izobraževalnih) zanimanj spletnih obiskovalcev.

- poizvedbe, ki so bile razvrščene v področje UDK 6 niso bile vsebinsko tesneje povezane s poizvedbami s področja UDK 5 kot sem spočetka predpostavil, iz česar lahko sklepam, da gre v tem primeru za dokaj nesorodne spletne obiskovalce in poizvedovalce.

- poizvedbe s področja UDK 1 in UDK 6 so si bile bolj sorodne (*npr. povezave: »stres -> transport, stres -> industrija, stres -> bolezen, shizofrenija -> bolezen, bolezensko stanje, psihologija, shizofrenija -> psihiatrija Etika -> Zdravstvo, Hipokratova zaprisega, Morala -> Zdravniki, Filozofija/psihologija -> Medicina, Temperamenti -> Hipokratovi kvadranti idr.*)

- presenetljivo je bilo opažanje, da sta si področje UDK 3 in UDK 5 precej blizu, kajti ugotovljena razdalja med obema področjima je bila (*kot že omenjeno*) zgolj 0,082792. V tem primeru lahko najdemo povezave med ključnimi besedami kot so npr. »Ekologija -> sociologija; naravna -> družbena« Še boljše povezave dobimo, če si ogledamo poizvedbe kot npr. Naravoslovje -> družbene vede; Narava -> družba, Statistika kaznivih dejanj -> matematika« V tem primeru gre po mojem mnenju za spletne obiskovalce, ki se zanimajo tako za družbene kot tudi za naravoslovne pojave z uporabo eksaktnih metod (*npr. statistika družbe -> matematična statistika*).

- najbolj sorodni so si bili spletni obiskovalci s področja UDK 9, UDK 8 in UDK 7, kajti spletni obiskovalci so mnogokrat poizvedovali po življenjepisih slavnih književnikov in skladateljev. Še zlasti zanimivi sta ključni besedi »Življenjepis in Biografija«. Ključna beseda »Življenjepis« je bila sorazmerno pogosta (*frekvenca enaka 39*), vendar je dosegla sorazmerno nizek rang in dejavnik **K** ( $r = 450$ ;  $K = 3,606$ ). Podobno lahko ugotovimo za ključno besedo »Biografija«, ki je prav tako bila sorazmerno pogosta (*frekvenca enaka 25*) in je dosegla nižji rang ( $r = 89$ ) in nekoliko nižjo vrednost **K** ( $K = 9,193$ ). Obe ključni besedi sta se večinoma pojavljali ob imenu in priimku znamenitih skladateljev in književnikov, medtem ko so se zgolj imena in priimki le-teh ob poizvedbah pojavljali pogosteje, zaradi česar je so npr. ključne besede kot so Bach, Vivaldi, Seliškar, Aškerc dosegle višji rang in višjo vrednost **K**.

## 6.2.6 Najpomembnejša izpeljana nova znanja/spoznanja

Gre za strnjeno predstavitev najpomembnejših izpeljanih novih znanj/spoznanj.

- 1.) Vedenje spletnih obiskovalcev/poizvedovalcev je sorazmerno predvidljivo, iz česar bi lahko sklepal, da imamo v večji meri opraviti s skupinami spletnih obiskovalcev, ki

izvedejo poizvedbe na podlagi določenega kolektivnega programa. Tovrstni ljudje so bolj ali manj stalni uporabniki spletnega UDK leksikona.

- 2.) Za proučevane spletne obiskovalce/poizvedovalce je značilen vedenjski vzorec, ki se kakovostno gledano nikoli ne spremeni in je sorazmerno neodvisen od časovne skale, tako da gre tudi v tem primeru za odkrit zakon moči. Spletni obiskovalci bodo v večini primerov za isti iskalni pojem poizvedovali enkrat, dvakrat, trikrat in morda po štirikrat, medtem ko bodo pogostejša poizvedovanja za določen iskalni pojem redkejša (*npr. še zlasti poizvedovanje za določen iskalni pojem po 10, 11, 15 in 22 krat je zelo redek pojav*).
- 3.) Ljudje v večini primerov izberejo načine lastnega vedenja, ki je pretežno v skladu s kolektivnimi (*družbenimi*) programi in njihovega pojma o sebi. Prav to znanje/spoznanje je možno prenesti na manjše množice ali skupine ljudi, znotraj katerih imajo posamezniki v sebi določena skladna oziroma podobna subjektivna (*podobne izkušnje*), objektivna (*podobna družbena znanja*) in interpersonalna znanja/spoznanja (*pridobljena znanja/spoznanja s pomočjo empatije in medosebnega sporazumevanja*). Posamezniki in s tem tudi (*homogene*) skupine gledajo na svet okoli sebe skozi prozorne obrazce, ki jih deloma ustvarjajo sami, a jih v večini primerov poskušajo prilagoditi resničnosti, iz katerih se svet sestoji. Po drugi strani pa posamezniki sprejmejo družbene prozorne obrazce, ki jih potem zaradi udobja nekoliko poenostavijo. Obnašanje skupin ima določeno stopnjo generalnosti, ki so v določeni meri tudi hierarhično organizirane in temu ustrezno strukturirane.
- 4.) Ključne besede kot so še zlasti npr. zgodovina, stres, Vivaldi, Haydn, Beethoven, Bach, Shakespeare in Napoleon imajo z vsebinskega vidika zelo širok spekter (*so interdisciplinarne in/ali multidisciplinarne narave*).
- 5.) Tudi v svetu znanja obstajajo določene generalnosti oziroma določeni miselni principi, ki so prav tako hierarhično organizirani in strukturirani, saj izhajajo iz človeka samega! Prav na podlagi poznavanja tovrstnih miselnih principov, je v naslednji stopnji možno izdelati učinkovite semantične mreže, ki nam kažejo povezave v obliki nadrejenosti/podrejenosti, pojmovne oddaljenosti, bližine, asociacije, polihierarhij idr.
- 6.) Pri segmentaciji spletnih obiskovalcev sem določil naslednje skupine in jim pripisal naslednje značilnosti:
  - a.) Pustolovci – ta skupina je kot že omenjeno največja. Tovrstni spletni obiskovalci/poizvedovalci so na notranjem iskalniku UDK leksikona izvedli največ poizvedb (*so bili najbolj dejavni*) in povrh tega so se njihove

uporabljene ključne besede uvrstile med najpomembnejše (*npr. zgodovina, Bach, Haydn, Napoleon*). Z vidika osnovnih miselnih silnic je za te spletne obiskovalce/poizvedovalce značilno, da pri poizvedovanju še zlasti izhajajo iz vsakanjega in filozofskega načina razmišljanja in redkeje iz libidnega načina razmišljanja. Globalno gledano se zanimajo še zlasti za tematike, ki so v povezavi z zabavo, igro, počutjem, hrano, počitnicami, družabnostjo, umetnostjo, znanostjo, poslovanjem in inovativnostjo. S tega je možno ugotoviti povezavo s psihološkimi vzgibi. Za tovrstne spletne obiskovalce/poizvedovalce so zelo značilni psihološki vzgibi kot npr. radovednost, veselje do lova, estetika/harmonija, humornost, so ambiciozni (*slo k uspehu*), tekmovalnost, veselje do športa in igre. Po drugi strani so nekoliko manj čustveni, lojalni oziroma zvesti in dajejo nekoliko manj poudarka na spolnost, obleko, hrano in pijačo. Z vidika UDK področij so tovrstni spletni obiskovalci/poizvedovalci iskali največkrat informacij s področja UDK 9, UDK 8 in UDK 7 (*gre za specializirane tematike s področja geografije in zgodovine držav, biografij znamenitih osebnosti in družin, pomembnih svetovnih zgodovinskih dogodkov, pomembnih književnih in glasbenih del itd.*).

- b.) Opazovalci – ta skupina je bila druga največja. Tovrstni spletni obiskovalci/poizvedovalci so bili zelo dejavni, vendar manj kot malo prej omenjeni. Njihove ključne besede so se na lestvici uvrstile nižje kot pa ključne besede, ki so jih uporabili »Pustolovci«, vendar pa višje kot pa ključne besede »Aplikativcev in Vsevedov«. Z vidika osnovnih miselnih silnic je za te spletne obiskovalce/poizvedovalce značilno, da pri poizvedovanju še zlasti izhajajo iz filozofskega načina razmišljanja, redkeje iz vsakdanjega načina razmišljanja in še redkeje iz libidnega načina razmišljanja. Globalno gledano se zanimajo še zlasti za tematike, ki so v povezavi z znanostjo, umetnostjo, s poslovanjem in inovativnostjo, manj se zanimajo za igre, počutje, prehrano, počitnice, družabnostjo in še manj za tematike v zvezi s spolnostjo, čustvi, ugodjem in občutki. Za tovrstne spletne obiskovalce/poizvedovalce so zelo značilni psihološki vzgibi kot npr. radovednost, vedoželjnost, estetika/harmonija, konstruktivnost, so ambiciozni (*slo k uspehu*), tekmovalnost, lojalnost/zvestoba in čistoča. Po drugi strani so nekoliko manj čustveni in dajejo malo poudarka na spolnost, čustev, obleko, hrano in pijačo. Z vidika UDK področij so

tovrstni spletni obiskovalci/poizvedovalci iskali največkrat informacij s področja UDK 5 in UDK 3 (*gre za specializirane tematike s področja matematike, statistike, ekologije, varstvo naravnega in družbenega okolja, fizike, migracije prebivalstva, natalitete in mortalitete prebivalstva itd.*).

- c.) Aplikativci – po velikosti in pomembnosti se je ta skupina spletnih obiskovalcev/poizvedovalcev uvrstila na tretje mesto. Z vidika osnovnih miselnih silnic je za te spletne obiskovalce/poizvedovalce značilno, da pri poizvedovanju enako močno izhajajo iz filozofskega, vsakdanjega in libidnega načina razmišljanja. Globalno gledano se zanimajo še zlasti za tematike, ki so v povezavi z znanostjo, umetnostjo, s poslovanjem in inovativnostjo, za igre, počutje, prehrano, počitnice, družabnostjo, spolnostjo, čustvi, ugodjem in občutki. Za tovrstne spletne obiskovalce/poizvedovalce so zelo značilni psihološki vzgibi kot npr. radovednost, vedoželjnost, konstruktivnost, ambicioznost (*slo k uspehu*), tekmovalnost, lojalnost/zvestoba in čistoča. Z razliko od prvih dveh skupin tovrstni spletni obiskovalci/poizvedovalci dajejo tudi velik poudarek na spolnost, čustva, obleko, hrano in pijačo. Z vidika UDK področij so tovrstni spletni obiskovalci/poizvedovalci iskali največkrat informacij s področja UDK 6 in UDK 1 (*gre za specializirane tematike s področja medicine dela, proizvodnje, organizacije proizvodnega dela, transporta, elektrotehnike, strojništva, gradbeništva, industrije, filozofije sistemov, filozofije življenja, psihologije dela, stresa itd.*).
- d.) Vsevedni – ta skupina je z glede na predhodne najbolj šibka in najmanj zastopana. Tovrstni spletni obiskovalci/poizvedovalci so bili z vidika poizvedovanja najmanj dejavni. Z vidika osnovnih miselnih silnic je za te spletne obiskovalce/poizvedovalce značilno, da pri poizvedovanju močno izhajajo iz filozofskega načina razmišljanja, manj iz vsakdanjega in še manj iz libidnega načina razmišljanja. Globalno gledano se zanimajo še zlasti za tematike, ki so v povezavi z znanostjo, umetnostjo, s poslovanjem in inovativnostjo, manj se zanimajo za igre, počutje, prehrano, počitnice, družabnost, spolnost, čustveni svet, ugodje in občutke. Za tovrstne spletne obiskovalce/poizvedovalce so zelo značilni psihološki vzgibi kot npr. vedoželjnost, konstruktivnost, ambicioznost (*slo k uspehu*), tekmovalnost, lojalnost/zvestoba in čistoča. Z razliko od prvih treh navedenih skupin tovrstni spletni obiskovalci/poizvedovalci dajejo zelo majhen poudarek na spolnost,

čustva, obleko, hrano in pijačo. Z vidika UDK področij so tovrstni spletni obiskovalci/poizvedovalci iskali največkrat informacij s področja UDK 0 in UDK 2 (*tudi ta naveza je v bistvu šibka – gre za specializirane tematike kot so npr. sistematika znanosti, metodologija, opredelitve, pismenstvo, informatika, računalništvo, bibliotekarstvo, muzeji, serijske publikacije, slovarje, leksikoni, kataloge, krščanska in nekrščanska verstva, mitologija, biblijske in mitološke osebe itd.*). Tovrstni spletni obiskovalci/poizvedovalci so največkrat iskali kratke opredlitve s spletnih leksikonov, slovarjev in enciklopedij.

Po moji oceni je možno tovrstno segmentacijo spletnih poizvedovalcev uporabiti ne samo v namen podpore vzgojnoizobraževalnega procesa, temveč tudi v smeri trženja tj. za npr. ugotavljanje določenih vedenjskih vzorcev potencialnih potrošnikov določenih proizvodov in storitev.

Na podlagi odkrivanja zanimivih vzorcev v podatkih in izpeljavi novih znanj/spoznanj sem ponovno potrdil prvo raziskovalno predpostavko: **»Potrdila se je hipoteza o univerzalnem informacijskem/poizvedovalnem vedenjskem vzorcu spletnih obiskovalcev glede (izobraževalnih) interesov (ožje gledano) ali širše gledano glede področnih interesov (npr. za poslovanje, za zdravje, za organizacijo dela, za izobraževanje, za zabavo itd. spletnih obiskovalcev), ki se tudi dolgoročno gledano ne spreminja in pri tem kot izid dobil še podrobnejše vzorce njihovega poizvedovalnega vedenja.«**

Prav zdaj lahko še za zaključek tega sklepnega poglavja predstavim izpeljano uporabno komponento vseh dosedanjih prizadevanj. Izid je bil kvalitativni model adaptivne digitalne knjižnice s pomembnim modulom MRT, ki ima naslednje prednosti in slabe strani:

Prednosti

- MRT uporabnikom omogoča uporabnikom bolj prilagojen in s tem posledično učinkovitejši dostop do uporabnih informacij (*zapisano velja tako za digitalne knjižnice na spletu kot tudi na CD-ROM*).
- da kot izid iskanja oziroma poizvedovanja daje nižjo stopnjo poizvedovalnega šuma
- da lahko MRT uporabimo kot orodje pri razreševanju informacijskih problemov tako v informacijsko dokumentalističnih centrih kot tudi v knjižnicah (*to v sklopu referenčnega procesa bibliotekar/dokumentalist v odnosu z uporabnikom*).
- da omogoča socialno zblíževanje z drugimi uporabniki, še zlasti sorodnimi

- da deluje kot odlična raziskovalna platforma glede raziskovanja uporabnikov in s tem posledično dodatno pripomore k večji prilagodljivosti informacijske ponudbe in storitev tako na področju bibliotekarskih kot tudi na področju pedagoških storitev!

#### Slabe strani

- MRT je lahko za neprofesionalne uporabnike manj razumljiv in je doba učenja daljša
- V MRT se lahko znotraj različnih dimenzij ponavljajo določene ključne besede, ki so lahko za neprofesionalne uporabnike zavajajoče

Model ADK z MRT sem tudi ovrednotil in kritično presodil, pri čemer sem prispel do naslednje ugotovitve in koristnega predloga za odpravo izpostavljenе pomanjkljivosti:

Na podlagi ovrednotenih usmeritev z ozirom na različne vrste uporabnikov sem ugotovil, da je dosegel kvalitativni model ADK z MRT zelo dobre izide. Največjo pomanjkljivost pa je možno zaznati prav pri usmeritvi »razumljivosti« (*ovrednotenje po vrsticah*), kajti če je določena aplikacija manj razumljiva za ključne uporabnike kot so pedagogi, študente in dijake, je tovrstna pomanjkljivost dokaj kritična. Glavni razlog za dozdevno nižjo stopnjo razumljivosti modela vidim na področju tezavra. Problem bi bilo možno rešiti tako, da bi izobraževalni zavodi uvedli nov študijski/učni predmet za učinkovito organizacijo in iskanje informacij (*usvajanje iskalnik tehnik, poznavanje koristnih zbirk podatkov, izdelava diagramov, poznavanje modelirnih orodij, izdelava digitalnega tezavra itd.*). Prav s tem obeležjem, bi bila še ta pomanjkljivost odpravljena in povrh tega bi tako pedagoški uslužbenci kot tudi študentje in dijaki pridobili koristna znanja, ki bodo še zlasti v prihodnosti (*morda*) postala nepogrešljiva!

## 7 Zaključek

Ob dolgoletnem zbiranju, opazovanju in površni anizi dejavnosti spletnih obiskovalcev na zasebni izvedbi spletnega UDK leksikona sem ob koncu pisanja tega doktorskega dela prišel do zaključka, da sem poprej mnogo ključnih dejstev spregledal. Zahvaljujoč spletometričnih meritev/analiz (*še zlasti metode zakona moči*) in uporabe določenih metod s področja OZVP sem se dokopal do zanimivih ugotovitev in v nadaljevanju do novih znanj in spoznanj. Bil sem prenesenečen in hkrati navdušen nad številnimi svetovi oziroma možnostmi, ki so se mi začele odpirati. Na koncu tega dela sem tudi razvil kvalitativni model ADK z modulom MRT in njegov prototip. Glavna zamisel tega dela je v bistvu bila vseskozi v tem, da sem poskušal dati poudarek na miselni hierarhiji uporabnikov (*spletni obiskovalci/poizvedovalci*), ki naj bi po mojem mnenju bili glavni smerokaz pri razvoju sodobnih računalniških aplikacij oziroma aplikacijskih sistemov.

### 7.1 Pogled v prihodnostni razvoj adaptivnih digitalnih knjižnic (ADK)

V prihodnosti bi morda lahko pričakovali inteligentne ADK (*okoli leta 2010 do 2015*), ki bi zmogli brati človekovo miselnost. Preizkusi v povezavi s »telepatijo« človeških misli s pomočjo vmesnika in računalnika že potekajo v Nemčiji in Združenih Državah Amerike. S posebnim vmesnikom (*ta je povezan z računalnikom*), ki je sestavljen iz številnih elektrod, ki se namestijo na roko in na glavo, se merijo odzivni vzorci testnih oseb na različne stimulatívne besede kot so npr. voda, sonce, zabava, strah, žalost in/ali vizualne prikaze s podobno vsebino. Računalnik tovrstne podatke procesira in jih potem prikaže v obliki vizualnih prikazov možganskih predelov, ki se glede na moč ali šibkost impulzov različno obarvajo. Prav zadnje navedeno je v bistvu napreden korak k temu, da dobijo sicer dokaj neoprijemljivi podatki kot so številke in besede s pomočjo metod eksaktnih znanosti neko oprijemljivejšo obliko in zgradbo.

# 8 Priloga

## 8.1 Slovar izrazov

Aplikacija je programska rešitev, ki kot sestavni del IS podpira neko logično celoto delovanja organizacijskega sistema

Aplikativni sistem je pojem, ki se ponavadi uporablja za bolj sestavljene rešitve, medtem ko se pojem aplikacija uporablja za manj sestavljene

Brskalnik (Browser) je program z grafičnim interaktivnim vmesnikom za iskanje, ogledovanje in upravljanje informacij v omrežju

Consistent Web node Framework diagram – diagram za prikazovanje spletnih povezav in spletnih vozlišč.

Digitalna knjižnica predstavlja zbirko medsebojno povezanih sistemov in virov, ki so uporabnikom dostopni, s pomočjo sodobne informacijske tehnologije. Specializirane kategorije digitalnih knjižnic so npr.: knjižnice na trdem disku – računalnik, knjižnice na CD-ROMu, knjižnice na spletu oziroma na internetu ipd. Ponekod še uporabljajo izraz virtualna knjižnica, kjer pa gre v bistvu za označbo zunanje oblike knjižnice kot tj. **KNJIŽNICA BREZ ZIDOV**. Adaptivne digitalne knjižnice so tiste, ki se prilagajajo uporabnikom.

E-izobraževanje je izobraževanje ljudi preko elektronskih medijev kot npr. s pomočjo elektronske pošte, s pomočjo interneta, multimedijev itd.

Informacijski Sistem (IS) je sistem, ki omogoča zbiranje, obdelavo, shranjevanje, distribucijo ter uporabo podatkov in informacij, še zlasti za odločanje v delovnih procesih

Informacijska tehnologija – označuje programsko, strojno in komunikacijsko opremo, ki se uporablja za podporo delovanja IS

Informacijski portal je predstavnik sodobne IT, ki uporabnikom omogoča lažji in hitrejši dostop do vsebin, za katere se zanimajo (za posebna področja npr. spletne knjižnice).

Internet je: 1. svetovno omrežje omrežij, ki uporablja protokole in storitve odprtih standardov, predvsem sklada protokolov TCP/IP. sin. medmrežje

2. globalni informacijski sistem

IP naslov (*IP address*) pomeni 32-bitni naslov Internetovega protokola, dodeljen gostitelju. IP naslov ima del, ki se nanaša na gostitelja, in del, ki se nanaša na omrežje

Iskalnik (spletni, inteligentni) **je aplikacija, ki služi kot orodje za iskanje informacij (na spletu).** Inteligentni iskalnik je orodje za inteligentno iskanje informacij (na spletu)

Odkrivanje zakonitosti v podatkih (*data mining*) je področje, ki temelji na metodah in algoritmih umetne inteligence in njenega ožjega področja strojnega učenja

Ontologija z vidika informatike je disciplina, ki se ukvarja z opredelitvijo besed in njihovim pomenom, kar se v nadaljnji stopnji uporablja za opis in predstavitev določenega znanja z bolj ali manj širokega področja.

Orodja za upravljanje z znanjem – predstavlja skupino različnih orodij za zbiranje, shranjevanje, analizo in dostop do uporabnih podatkov oziroma informacij, ki omogočajo hitre in kakovostne odločitve (*informacije so že pripravljene in mnogokrat temeljijo na bogatih izkušnjah različnih ljudi*)

RDF (*Resource Description Framework*) je standard za opredelitev metapodatkov oziroma opis spletnih virov RDF, ki se lahko uporablja za iskanje virov, za katalogizacijo – opis vsebine, opisovanje pravic intelektualne lastnine itd.

Spletometrija (*webometrija*) merjenje in ovrednotenje pojavov, procesov in izdelkov na spletu (širše gledano: na internetu – *Cybermetrics*)

Stran (*page*) je hipermedijski dokument v Webu (*omrežju*)

Taksonomija pomeni razvrstitev prvin v določen klasifikacijski sistem, v katerem so tudi bolj ali manj natančno ponazorjene določene povezave

TCP/IP (*Transmission Control Protocol/Internet Protocol*) je protokol za prenos podatkov

UDK (*Universal Decimal Classification*) – univerzalna decimalna klasifikacija je klasifikacijski sistem, ki razvršča vso človekovo znanje s pomočjo vrstilcev (števil) in drugih dodatnih znamenj

Tezaver je standardizirani referenčni slovar, ki je zbirka besed naravnega jezika

(*splošnega in strokovnega*) s predstavitvijo njihovih pojmovnih odnosov in nam služi kot pomoč pri dokumentaciji. Multidimenzionalni rang tezaver je tezaver, ki večnamensko vključuje več različnih dimenzij.

UML – *Unified Modeling Language* – enotni slikovni jezik za objektno modeliranje

Univerzalni lokator virov (*URL - Uniform Resource Locator*) je oblika naslova mesta, ki razkriva ime strežnika, v katerem so shranjene datoteke mesta, pot do datoteke in ime datoteke

**XML (*Extensible Markup Language*) je definicija metod za shranjevanje in obdelavo besedil v elektronski obliki, ki je neodvisna od konkretnih računalniških naprav in sistemov. XML je poenostavljena različica SGML**

**Web Browser je brskalnik za svetovni splet**

**Zakon moči (Power Law) – metoda, s pomočjo katere se ugotavlja neko univerzalno nagnjenje, ki je sorazmerno neodvisno od časa. Med tovrstne zakone prištejemo Zipfov zakon in fizikalne različice tega zakona, Bradfordov zakon, Lotkov zakon idr.**

## 9 Viri in uporabljena programska orodja

- 1.) Ackermann, E. & Hartmann, K.(2000). The information specialist's guide to searching & researching on the Internet & the World Wide Web. 2 nd. ed. Chicago, London : Fitzroy Dearborn Publishers, 438 str.
- 2.) Aitchison, J.(2000). Thesaurus construction and use : a practical manual. London: Aslib IMI, XIV, 218 str.
- 3.) Arms, W.Y.(2000). Digital Libraries. Cambridge (Mass.) ; London : The MIT Press. X, 287 str. (Digital libraries and electronic publishing)
- 4.) Bernstein, P.(1999). Finding statistics online : how to locate the elusive numbers you need. Medford : Information Today, 356 str.
- 5.) Berry, M. J. A. & Linoff, G.S.(2004). Data mining techniques : for marketing, sales, and customer relationshipmanagement. - 2nd ed. - Indianapolis (Ind.) : Wiley Pub. - XXV, 643 str.
- 6.) Björneborn, L.(2004). Small-world link structures across an academic web space : a library and information science approach. Forlag: Department of Information Studies, Royal School of Library and Information Science, 399 str.
- 7.) Broughton, V.(2006). Essential Thesaurus construction. London: Facet, 296 str.
- 8.) Chowdhury, G.G.(2003). Introduction to digital libraries. London : Facet. XV, 359 str.
- 9.) DE Ville, B.(2001.) Microsoft data mining : integrated business intelligence for e-commerce and knowledge management/Barry de Ville. - Boston [etc.] : Digital Press. - XX, 315 str.
- 10.) Ghaoui, C.(2004). E-education applications : human factors and innovative approaches. Hershey (PA), London [etc.] : Information Science Publishing : Idea Group, 363 str.
- 11.) KANTARDZIC, M.(2003). Data mining : concepts, models, methods, and algorithms. Hoboken (NJ) : Wiley-Interscience : IEEE Press, XII, 345 str.
- 12.) Kljajić, M.(2002). Teorija sistemov. Kranj: Moderna organizacija, 238 str.
- 13.) Konchady, M.(2006). Text Mining Application Programming. Boston: Charles River Media. XIX, 412 str. + CD-ROM.
- 14.) LYKKE Nielsen, M.(2002). The word association method : a gateway to work-task based retrieval. Abo : Abo Akademis förlag, VIII., 342 str.

- 15.) MACHINE learning and data mining : methods and applications/edited by Ryszard S. Michalski, Ivan Bratko, Miroslav Kubat. - Chichester [etc.] : J. Wiley & Sons, cop. 1998. - XVI, 456 str.
- 16.) Martin, E.(2004). Specialist computing's computer jargon : dictionary and thesaurus. Tonbridge: Biddles, 352 str.
- 17.) MAIMON, Oded Z.(2001). Knowledge discovery and data mining : the info-fuzzy network (IFN) methodology. Dordrecht ; Boston ; London : Kluwer Academic, IX, 171 str. (Massive computing)
- 18.) Mayr, P.(2004). Entwicklung und Test einer logfilebasierten Metrik zur Analyse von Website entries am Beispiel einer akademischen Universitäts-Website. Berlin : Inst. für Bibliothekswiss, 106 str. (Humboldt-Universität zu Berlin, Institut für Bibliothekswissenschaft).
- 19.) McCue, C.(2007). Data Mining and Predictive Analyses : intelligence gathering and crime analyses. Oxford ...: Butterworth-Heinemann. XXXI, 332 str.
- 20.) Mena, J.(2003). Investigative Data Mining for Security and Criminal Detection. Oxford: Buterworth-Heinemann. XVI, 452 str.
- 21.) Pace, A. K.(2003). The ultimate digital library : where the new information players meet. Chicago : American Library Association. XVII, 168 str.
- 22.) Urbanija, J.(1996). Metodologija izdelave tezavra. Ljubljana: Filozofska fakulteta, Oddelek za bibliotekarstvo, 70 str.
- 23.) WESTPHAL, C. & Blaxton, T.(1998). Data mining solutions : methods and tools for solving real-world problems. New York [etc.] : J. Wiley & Sons, XXII, 617 str. :
- 24.) WITTEN, Ian H. (2000). Data mining : practical machine learning tools and techniques with Java implementations/Ian H. Witten, Eibe Frank. - San Francisco (California) : Morgan Kaufmann. - XXV, 371 str. : ilustr. ; 24 cm. - (The Morgan Kaufmann series in data management systems)

### **Science Citation Index (SCI) članki**

- 1.) Björneborn, L. & Ingversen, P.(2001). Perspectives of webometrics. *Scientometrics*, 50(1): 65-82.
- 2.) Björneborn, L. & Ingversen, P.(2004). Toward a basic framework for webometrics. *Journal of the American Society for Information Science and Technology*, 55(14): 1216-1227.

- 3.) Jespen, E.T., Seiden, P., Ingversen, P. & Björneborn, L.(2004). Characteristics of scientific web publications: preliminary data gathering and analysis. *Journal of the American Society for Information Science and Technology*, 55(14): 1239 - 1249.
- 4.) Kochtanek TR, Hein KK & Kassim ARC.(2001). A digital library resource Web site: Project DL. *ONLINE INFORM REV*, 25 (1), 29-40.
- 5.) Koohang, A.(2004). Students' perceptions toward the use of the digital library in weekly web-based distance learning assignments portion of a hybrid programme. *BRITISH JOURNAL OF EDUCATIONAL TECHNOLOGY*, 35 (5), 617-626.
- 6.) Lamirel, J.C., Francois, C., Al Shehabi, S. & Hoffmann, M.(2004). New classification quality estimators for analysis of documentary information: Application to patent analysis and web mapping. *Journal of the American Society for Information Science and Technology*, 60(3): 445 - 462.
- 7.) McKay D. & Cunningham S.J.(2003). Browsing a Digital Library : A new approach for the New Zealand Digital Library. *DIGITAL LIBRARIES: TECHNOLOGY AND MANAGEMENT OF INDIGENOUS KNOWLEDGE FOR GLOBAL ACCESS LECTURE NOTES IN COMPUTER SCIENCE*, 2911, 329-339.
- 8.) PIVK, A., CIMIANO, P., SURE, Y., GAMS, M., RAJKOVIČ, V. & STUDER, R.(2007) Transforming arbitrary tables into logical form with TARTAR. *Data knowl. eng.* [Print ed.], vol. 60, str. 567-595.
- 9.) Rosemann, U.(2003). Information networks and specialized virtual libraries: The beginning of a marvelous collaborative effort?. *ZEITSCHRIFT FÜR BIBLIOTHEKSWESEN UND BIBLIOGRAPHIE*, 50 (1), 13-18 JAN-FEB.
- 10.) Saracevic T. & Dalbello, M.(2001). A survey of digital library education. *PROCEEDINGS OF THE 64TH ASIST ANNUAL MEETING, VOL 38, PROCEEDINGS OF THE ASIST ANNUAL MEETING*, 38, 209-223.
- 11.) Thelwall, M. & Vaughan, L.(2004). Webometrics: An introduction to the special issue. *Journal of the American Society for Information Science and Technology*, 55(14): 1213 - 1215.
- 12.) Thelwall, M., Vaughan, L. & Björneborn, L.(2005). Webometrics. *Annual Review of Information Science and Technology*, 39: 81 - 135.

## Drugi članki

- 1.) Bartol, T.(2001). Mapping of contents in "Zbornik Biotehniške fakultete Univerze v Ljubljani. V: Zbornik Biotehniške fakultete Univerze v Ljubljani. Kmetijstvo. Zootehnika. Agricultural issue. Zootechnica. - ISSN 1408-3434. Letn.
- 2.) BELLAZZI, R. & Zupan, B.(2001). Intelligent data analysis - special issue. V: Methods of information in medicine. - ISSN 0026-1270. Vol. 40, no. 5 (2001), str. 362-364.
- 3.) BRATKO, I. (2004). Qualitative data mining/Ivan Bratko, Dorian Suc. - Bibliografija: str. 12. V: The new trends in knowledge processing data mining, semantic web and computational science/[organized by The Institute of Scientific and Industrial Research (SANKEN)]. - [S. l. : s. n.], 2003. - Str. 10-12.
- 4.) Buxton, A. B.(1993). "Computer searching on UDC numbers". Encyclopedia of Library and Information Science, 52,132-151.
- 5.) Curk T., Demsar J., Xu Q., Leban G., Petrovic U., Bratko, I., Shaulsky G. & Zupan, B. Microarray data mining with visual programming. *Bioinformatics* 21(3): 396-398.
- 6.) Gerlič, I. & Rajkovič, V.(2005). Didaktični vidiki uporabe informacijske in komunikacijske tehnologije - IKT : (poučevanje in učenje). Uporaba računalnika pri pouku tehnike in tehnologije, Maribor : Univerza v Mariboru, Pedagoška fakulteta, Znanstveni inštitut in Center za računalništvo, informatiko in multimedije v izobraževanju, 30 f.
- 7.) Lavbič, D. & Krisper, M.(2005). Semantika podatkov in ontologije. V: Uporabna informatika. - ISSN 1318-1882. Letn.13, št. 3 (2005), str. 121-135.
- 8.) Krisper, M.& Rupnik, R.(2004). Aplikativni sistemi za odkrivanje zakonitosti v podatkih
- 9.) POGAČNIK, M. & Tasič, J. F.(2000). Data mining on the Internet. V: Zbornik devete Elektrotehniške in računalniške konference ERK, 21. - 23. september 2000, Portorož, Slovenija. Ljubljana : IEEE Region 8, Slovenska sekcija IEEE. - ISBN 961-6062-19-0. - zv. A, str. 87-90.
- 10.) Pivec, M. & Rajkovič, V.(1999). Upravljanje znanja znotraj učnih okolij. V: Uporabna informatika. - Let. 7, št. 4, str. 42-46.
- 11.) SADIKOV, A., BRATKO, I. & KONONENKO, I. Search versus knowledge : an empirical study of minimax on KRK. V: HERIK, Jaap van den (ur.), IIDA, Hiroyuki (ur.), HEINZ, Ernst A. (ur.). Advances in computer games : many games, many challenges : proceedings of the ICGA/IFIP SG16, (IFIP, 135). Boston; Dordrecht; London: Kluwer Academic Publishers, str. [33]-44 - PDF.

12.) Woldering, B.(2004). EU-project The European Library (Tel) successfully finished.  
Knjižnica, 48(1-2), 223-235

#### **Glavni spletni (medmrežni) viri**

- 1.) Chen, H., Chau, M., and Zeng, D., "[CI Spider: A Tool for Competitive Intelligence on the Web](#)," *Decision Support Systems (DSS)*, 34(1), 1-17, 2002.
- 2.) Chen, H., Fan, H., Chau, M., and Zeng, D., "[MetaSpider: Meta-Searching and Categorization on the Web](#)," *Journal of the American Society for Information Science and Technology (JASIST)*, 52(13), 1134-1147, 2001.
- 3.) Chau, M., Fang, X., and Sheng, O. R. L., "[Analysis of the Query Logs of a Web Site Search Engine](#)," *Journal of the American Society for Information Science and Technology (JASIST)*, 56(13), 1363-1376, 2005.
- 4.) Chau, M., Chen, H., Qin, J., Zhou, Y., Qin, Y., Sung, W. K., and McDonald, D., "[Comparison of Two Approaches to Building a Vertical Search Tool: A Case Study in the Nanotechnology Domain](#)," in *Proceedings of The Second ACM/IEEE-CS Joint Conference on Digital Libraries (JCDL'02)*, Portland, Oregon, USA, July 14-18, 2002, pp. 135-144.

#### **Digitalne knjižnice (DK) Evropska unija (EU)**

[http://www.delos.info/files/pdf/events/2004\\_Jul\\_8\\_10/D8.pdf](http://www.delos.info/files/pdf/events/2004_Jul_8_10/D8.pdf) (2007-06-02) Prihodnja vizija DK

[http://www.ercim.org/publication/Ercim\\_News/enw66/ioannidis.html](http://www.ercim.org/publication/Ercim_News/enw66/ioannidis.html) (2007-06-02)

Prihodnost DK in dosedanje stanje

<http://delos-noe.iei.pi.cnr.it/activities/researchforum/Brainstorming/brainstorming-report.pdf>  
(2007-06-02) EU raziskovalni program DK

[http://ec.europa.eu/information\\_society/activities/digital\\_libraries/index\\_en.htm](http://ec.europa.eu/information_society/activities/digital_libraries/index_en.htm) (2007-06-02)

Digitalne knjižnice v EU leta 2010

<http://www.davidmattison.ca/wordpress/?p=1416> (2007-06-02) Vizija EU glede DK

[http://www.edis.sk/ekes/interim\\_report\\_16\\_10\\_06.pdf](http://www.edis.sk/ekes/interim_report_16_10_06.pdf) (2007-06-02) Pobude EU glede DK

[http://ec.europa.eu/research/science-society/document\\_library/pdf\\_06/communication-022007\\_en.pdf](http://ec.europa.eu/research/science-society/document_library/pdf_06/communication-022007_en.pdf) (2007-06-02)

#### **Pomožni spletni (medmrežni) viri**

- 1.) <http://www.crisp-dm.org/CRISPWP-0800.pdf> (2006-06-12)

2.)

[http://www.siggraph.org/education/materials/HyperVis/applicat/data\\_mining/data\\_mining.html](http://www.siggraph.org/education/materials/HyperVis/applicat/data_mining/data_mining.html) (2006-07-21)

3.) <http://www.thearling.com/text/scoring/scoring.htm> (2006-07-14)

4.) <http://www.webology.ir/2004/v1n2/a6.html> (2006-07-14)

5.) [http://www.insuma.de/insuma/de/SEspell\\_demo.html](http://www.insuma.de/insuma/de/SEspell_demo.html) (2006-02-12)

6.) <http://www.searchtools.com/guide/index.html> (2006-05-14)

7.) <http://de.wikipedia.org/wiki/Suchmaschine> <http://german.therfcc.org/webometrie-413449.html> [http://es.csiro.au/pubs/melucci\\_hawking.pdf](http://es.csiro.au/pubs/melucci_hawking.pdf) (2006-05-14)

8.) <http://polaris.gseis.ucla.edu/jfurner/jasist03.pdf> (2006-05-14)

9.) [http://individual.utoronto.ca/azuccala\\_web/Zuccala-WebIntelligence.pdf](http://individual.utoronto.ca/azuccala_web/Zuccala-WebIntelligence.pdf) (2006-07-05)

10.) [http://www.citi.pt/monografias/estudos/search\\_engines.doc](http://www.citi.pt/monografias/estudos/search_engines.doc) (2006-07-05)

11.) <http://informationr.net/ir/6-1/paper90.html> (2006-05-12)

12.) <http://pespmc1.vub.ac.be/SUPORGLI.html> (2006-05-12)

13.) <http://www.business.hku.hk/~mchau/publications.html> (2006-05-12)

14.) [http://home.izum.si/cobiss/cobiss\\_obvestila/2000\\_3/Html/clanek\\_01.html](http://home.izum.si/cobiss/cobiss_obvestila/2000_3/Html/clanek_01.html) (2005-03-12)

15.) <http://www.ib.hu-berlin.de/~kumlau/handreichungen/h67> (2005-03-10)

16.) <http://www.gragert.de/magister/motivation.html> (2005-03-10)

17.) [http://www.vww-info.de/Magisterarbeit\\_Bettina\\_Wilhelm.pdf](http://www.vww-info.de/Magisterarbeit_Bettina_Wilhelm.pdf) (2005-03-10)

18.) <http://www.searchengineshowdown.com/stats> (2005-03-10)

19.) <http://searchenginewatch.com/reports/article.php/2156461> (2005-03-10)

20.) [http://www.gerhard.de/gerold/owa/gerhard.create\\_index\\_html?form\\_language=99](http://www.gerhard.de/gerold/owa/gerhard.create_index_html?form_language=99)  
(2005-03-10)

21.) <http://www.1cog.com/search-engine-statistics.html> (2005-03-10)

22.) <http://internet-statistics-guide.netfirms.com> (2005-03-10)

23.) [http://www.w3schools.com/browsers/browsers\\_stats.asp](http://www.w3schools.com/browsers/browsers_stats.asp) (2005-03-10)

24.) <http://www.why-not.com/company/stats.htm> (2005-03-10)

25.) <http://www.internetworldstats.com/stats.htm> (2005-03-10)

26.) <http://www.mit.edu/people/mkgray/growth> (2005-03-10)

27.) <http://www.nla.gov.au/oz/stats.html> (2005-03-10)

28.) <http://www.helsinki.fi/~jpuranen/links.html> (2005-03-10)

29.) <http://www.xtremes.math.uni-siegen.de/reiss/research.htm> (2005-03-10)

30.) <http://english.unitecology.ac.nz/resources/resources/learntolearn/module1.html>  
(2004-07-10)

- 31.) <http://www.ciolek.com/WWWVL-InfoQuality.html> (2004-07-25)  
32.) <http://www2.library.unr.edu/nla2003/virtuallib.html> (2004-07-08)

**9.1 Priporočeni viri na podlagi posebnih izvedenih poizvedb s ključno besedo  
»MULTIDIMENSIONAL THESAURUS« na Web of Science, INSPEC in ERIC  
Web of Science – DVA ZADETKA**

**Record 1 of 2**

**Author(s):** Park, H

**Title:** Relevance of science information: Origins and dimensions of relevance and their implications to information retrieval

**Source:** INFORMATION PROCESSING & MANAGEMENT, 33 (3): 339-352 MAY 1997

---

**Record 2 of 2**

**Author(s):** Lerat, A

**Title:** On the analysis of implicit schemes of the Lax-Wendroff type for an arbitrary number of space dimensions.

**Source:** RECHERCHE AEROSPATIALE, (6): 373-387 1995

**INSPEC – OSEM ZADEKOV**

1. **A cooccurrence-based thesaurus and two applications to information retrieval**  
[Schutze, H.](#) (Xerox Palo Alto Res. Center, CA, USA); [Pedersen, J.O.](#) **Source:** *Information Processing & Management*, v 33, n 3, May 1997, p 307-18  
**Database:** Inspec  
[Abstract](#) - [Detailed](#) - [Full-text](#)
2. **Thesaurus and ontology, or formalization of the generic inheritance of conceptual features**  
[Ungvary, R.](#) (Orszagos Szechenyi Konyvtar, Hungary) **Source:** *Tudomanyos es Muszaki Tajekoztatas*, v 51, n 5, 2004, p 175-91 **Language:** Hungarian  
**Database:** Inspec  
[Abstract](#) - [Detailed](#)
3. **A multi-facet taxonomy system with applications in unstructured knowledge management**  
[Cheung, C.F.](#); [Lee, W.B.](#); [Wang, Y.](#) **Source:** *Journal of Knowledge Management*, v 9, n 6, 2005, p

76-91

**Database:** Inspec

[Abstract](#) - [Detailed](#) - [Full-text](#)

- 4. **An alternative perspective on medical informatics structure: EMBASE vs. MEDLINE**  
[Morris, T.A.](#) (Sch. of Libr. & Inf. Sci., Kent State Univ., Columbus, OH, USA) **Source:** *Proceedings of the 66th ASIST Annual Meeting. Humanizing Information Technology: From Ideas to Bits and Back. ASIST 2003*, 2003, p 528  
**Database:** Inspec  
[Abstract](#) - [Detailed](#)
  
- 5. **Toward semantic-based retrieval of visual information: a model-based approach**  
[Youngchoon Park](#) (Dept. of Comput. Sci., Arizona State Univ., Tempe, AZ, USA); [Golshani, F.](#); [Panchanathan, S.](#) **Source:** *Proceedings of the SPIE - The International Society for Optical Engineering*, v 4862, 2002, p 50-61  
**Database:** Inspec  
[Abstract](#) - [Detailed](#) - [Full-text](#)
  
- 6. **Adaptive visual regions categorization with sets of points of interest**  
[Houissa, H.](#) (INRIA-Rocquencourt, Le Chesnay, France); [Boujemaa, N.](#); [Frigui, H.](#) **Source:** *Advances in Multimedia Information Processing - PCM 2006 7th Pacific Rim Conference on Multimedia. Proceedings (Lecture Notes in Computer Science Vol. 4261)*, 2006, p 485-93  
**Database:** Inspec  
[Abstract](#) - [Detailed](#)
  
- 7. **On the drafting of classification media for the documentation of radiologic diagnosis by means of SOPS AIDOS**  
[Schuler, W.](#) (Bezirkskrankenhaus, Radiologische Klinik und Poliklinik, Karl-Marx-Stadt, East Germany); [Kunzelmann, F.](#); [Kretzschmar, U.](#) **Source:** *Informatik*, v 25, n 4, 1978, p 39-42 **Language:** German  
**Database:** Inspec  
[Abstract](#) - [Detailed](#)
  
- 8. **Relevance of science information: origins and dimensions of relevance and their implications to information retrieval**  
[Hongseok Park](#) **Source:** *Information Processing & Management*, v 33, n 3, May 1997, p 339-52

**Database:** Inspec

[Abstract](#) - [Detailed](#) - [Full-text](#)

**ERIC – NIČ ZAETKOV**

**LISA – NIČ ZAETKOV**

## **9.2 Priporočeni viri s področja adaptivnih digitalnih knjižnic z zbirki podatkov Web of Science (WOS), ERIC, LISA, Compendex in INSPEC od leta 2000 do 2007**

Selected Records

1 - 25 of 25 selected records

- [Remove](#) 1. **Data for the future: The German project "Co-operative development of a long-term digital information archive" (kopal)**

[Altenhoner, R.](#) (Die Deutsche Bibliothek, Frankfurt Am Main, Germany) **Source:** *Library Hi Tech*, v 24, n 4, 2006, p 574-82

**Database:** Inspec

[Full-text](#)

- [Remove](#) 2. **Towards a self healing information system for digital libraries**

[Ambati Vamshi](#) (Inst. for Software Res. Int., Carnegie Mellon Univ., Pittsburgh, PA, USA); [Reddy Raj](#) **Source:** *Journal of Zhejiang University (Science)*, v 6A, n 11, Nov. 2005, p 1221-8

**Database:** Inspec

[Full-text](#)

- [Remove](#) 3. **Hardware-driven adaptive  $\kappa$ -means clustering for real-time video imaging**

[Maliatski, Boris](#) (VLSI Systems Center, Ben-Gurion University); [Yadid-Pecht, Orly](#) **Source:** *IEEE Transactions on Circuits and Systems for Video Technology*, v 15, n 1, January, 2005, p 164-166

**Database:** Compendex

[Full-text](#)

- [Remove](#) 4. **Creating metadata that work for digital libraries and Google**

[Dawson, A.](#) (Dept. of Comput. & Inf. Sci., Univ. of Strathclyde, Glasgow, UK) **Source:** *Library Review*, v 53, n 7, 2004, p 347-50

**Database:** Inspec

[Full-text](#)

[Remove](#)

5. **Digital identity matters**

[Allison, A.](#) (Humanities Adv. Technol. & Inf. Inst., Glasgow Univ., UK); [Currall, J.](#); [Moss, M.](#); [Stuart, S.](#) **Source:** *Journal of the American Society for Information Science and Technology*, v 56, n 4, 2005, p 364-72

**Database:** Inspec

[Full-text](#)

[Remove](#)

6. **Automated user modeling for personalized digital libraries**

[Frias-Martinez, E.](#) (Department of Information Systems and Computing, Brunel University); [Magoulas, G.](#); [Chen, S.](#); [Macredie, R.](#) **Source:** *International Journal of Information Management*, v 26, n 3, June, 2006, p 234-248

**Database:** Compendex

[Full-text](#)

[Remove](#)

7. **Personalized and collaborative digital library capabilities: Responding to the changing nature of scientific research**

[Luce, Rick](#) (Los Alamos National Laboratory); [Di Giacomo, Mariella](#) **Source:** *Science and Technology Libraries*, v 24, n 1-2, 2003, p 135-152

**Database:** Compendex

[Full-text](#)

[Remove](#)

8. **InfoWeb: an adaptive information filtering system for the cultural heritage domain**

[Gentili, G.](#) (AI Lab., STE SpA, Pomezia, Italy); [Micarelli, A.](#); [Sciarrone, F.](#) **Source:** *Applied Artificial Intelligence*, v 17, n 8-9, Sept.-Oct. 2003, p 715-44

**Database:** Inspec

[Full-text](#)

[Remove](#)

9. **Lossless compression of large binary images in digital spatial libraries**

[Ageenko, E.](#) (Dept. of Comput. Sci., Joensuu Univ., Finland); [Franti, P.](#) **Source:** *Computers & Graphics*, v 24, n 1, Feb. 2000, p 91-8

**Database:** Inspec

[Full-text](#)

[Remove](#)

10. **Learning lessons holistically in the Glasgow Digital Library**

[Nicholson, D.](#) (Strathclyde Univ., Glasgow, UK); [Macgregor, G.](#) **Source:** *D-Lib Magazine*, v 8, n 7-8, July-Aug. 2002

**Database:** Inspec

[Full-text](#)

[Remove](#) 11. **Automated user modeling for personalized digital libraries**

[Chen, S.](#) (Dept. of Inf. Syst. & Comput., Brunel Univ., Middlesex, UK); [Frias-Martinez, E.](#); [Magoulas, G.](#); [Macredie, R.](#) **Source:** *International Journal of Information Management*, v 26, n 3, June 2006, p 234-48

**Database:** Inspec

[Full-text](#)

[Remove](#) 12. **User interaction in digital libraries: coping with diversity through adaptation**

[Stephanidis, C.](#) (Inst. of Comput. Sci., Hellas-Found. for Res. & Technol., Heraklion, Greece); [Akoumianakis, D.](#); [Paramythis, A.](#); [Nikolaou, C.](#) **Source:** *International Journal on Digital Libraries*, v 3, n 2, Aug. 2000, p 185-205

**Database:** Inspec

[Full-text](#)

[Remove](#) 13. **Adaptive load sharing for clustered digital library servers**

[Zhu, H.](#) (Dept. of Comput. Sci., California Univ., Santa Barbara, CA, USA); [Yang, T.](#); [Zheng, Q.](#); [Watson, D.](#); [Ibarra, O.H.](#); [Smith, T.](#) **Source:** *International Journal on Digital Libraries*, v 2, n 4, May 2000, p 225-35

**Database:** Inspec

[Full-text](#)

[Remove](#) 14. **A framework for performance monitoring, load balancing, adaptive timeouts and quality of service in digital libraries**

[Kapidakis, S.](#) (Inst. of Comput. Sci., FORTH, Heraklion, Greece); [Terzis, S.](#); [Sairamesh, J.](#) **Source:** *International Journal on Digital Libraries*, v 3, n 1, July 2000, p 19-35

**Database:** Inspec

[Full-text](#)

[Remove](#) 15. **Emerging tools for evaluating digital library services: Conceptual adaptations of LibQUAL+ and CAPM**

[Heath, F.](#) (Texas A&M Univ., College Station, TX, USA); [Kyrillidou, M.](#); [Webster, D.](#)

[Choudhury, S.](#); [Hobbs, B.](#); [Lorie, M.](#); [Flores, N.](#) **Source:** *JoDI - Journal of Digital Information*, v 4, n 2, 2003

**Database:** Inspec

**Remove** 16. **A practical consistent-quality two-pass VBR video coding algorithm for digital storage application**

[Hai Bing Yin](#) (Inst. of Image Commun. & Inf. Process., Shanghai Jiaotong Univ., China); [Xiang Zhong Fang](#); [Li Chen](#); [Jun Hou](#) **Source:** *IEEE Transactions on Consumer Electronics*, v 50, n 4, Nov. 2004, p 1142-50

**Database:** Inspec

**Full-text**

**Remove** 17. **InfoWeb: An adaptive information filtering system for the cultural heritage domain**

[Gentili, Gianluigi](#) (Universita RomaTre, Dipto. di Informatica e Automazione, AI Lab); [Micarelli, Alessandro](#); [Sciarrone, Filippo](#) **Source:** *Applied Artificial Intelligence*, v 17, n 8-9, September/October, 2003, p 715-744

**Database:** Compendex

**Full-text**

**Remove** 18. **Font adaptive word indexing of modern printed documents**

[Marinai, S.](#) (Dipt. di Sistemi e Inf., Univ. di Firenze, Italy); [Marino, E.](#); [Soda, G.](#) **Source:** *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v 28, n 8, Aug. 2006, p 1187-99

**Database:** Inspec

**Full-text**

**Remove** 19. **Optimal design of adaptive equalizer based on DVB-C**

[Huang Wei](#) (Eng. & Technol. Res. Centre, Southeast Univ., Nanjing, China); [Wu Jian-hui](#); [Yan Fei](#); [Liu Xin](#); [Zhang Meng](#) **Source:** *Video Engineering*, n 7, 2006, p 31-4 **Language:** Chinese

**Database:** Inspec

**Remove** 20. **Adaptive functional programming**

[Acar, Umut A.](#) (Toyota Technological Institute); [Blleloch, Guy E.](#); [Harper, Robert](#) **Source:** *ACM Transactions on Programming Languages and Systems*, v 28, n 6, 2006, p 990-1034

**Database:** Compendex

**Full-text**

**Remove** 21. **Dynamic pipeline design of an adaptive binary arithmetic coder**

[Kuang, S.R.](#) (Dept. of Electrical Engineering, National Cheng Kung University); [Jou, J.M.](#); [Chen, R.D.](#); [Shiau, Y.H.](#) **Source:** *IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing*, v 48, n 9, September, 2001, p 813-825

**Database:** Compendex

[Full-text](#)

**Remove** 22. **An adaptive video multicast scheme for varying workloads**

[Hua, Kien A.](#) (Computer Science Program, School of EECS, University of Central Florida); [Oh, JungHwan](#); [Vu, Khanh](#) **Source:** *Multimedia Systems*, v 8, n 4, 2002, p 258-269

**Database:** Compendex

[Full-text](#)

**Remove** 23. **Using metadata to integrate digital libraries by three-layer architecture**

[Huang, Su-Hsien](#) (Institute of Computer Science and Engineering, National Chiao Tung University); [Ke, Hao-Ren](#); [Yang, Wei-Pang](#) **Source:** *WSEAS Transactions on Computers*, v 5, n 10, October, 2006, p 2301-2308

**Database:** Compendex

**Remove** 24. **Dynamic pipeline design of an adaptive binary arithmetic coder**

[Shiann Rong Kuang](#) (Dept. of Electron. Eng., Southern Taiwan Univ. of Technol., Tainan, Taiwan); [Jer Min Jou](#); [Ren Der Chen](#); [Yeu Horng Shiau](#) **Source:** *IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing*, v 48, n 9, Sept. 2001, p 813-25

**Database:** Inspec

[Full-text](#)

**Remove** 25. **An adaptive video multicast scheme for varying workloads**

[Hua, K.A.](#) (Comput. Sci. Program, Univ. of Central Florida, Orlando, FL, USA); [JungHwan Oh](#); [Vu, K.](#) **Source:** *Multimedia Systems*, v 8, n 4, 2002, p 258-69

**Database:** Inspec

[Full-text](#)

1. **Contractual and technological approaches for protecting digital works: Their relationship with copyright limitations**

[Fernandez-Molina, J. Carlos](#) (Dept. of International Relations, Sch. of Library Information Science, University of Granada) **Source:** *Online Information Review*, v 28, n 2, 2004, p 148-157

**Database:** Compendex

[Abstract](#) - [Detailed](#) - [Full-text](#)

- 2. **Documenting the Archive - Using content analysis techniques**  
[Messina, Alberto](#) (RAI CRIT) **Source:** *EBU Technical Review*, n 305, January, 2006, p 69-78  
**Database:** Compendex  
[Abstract](#) - [Detailed](#)
  
- 3. **Looking for good art web resources and image databases part three general and U.S. tools**  
[Mattison, David](#) **Source:** *Searcher:Magazine for Database Professionals*, v 12, n 10, November/December, 2004, p 24-32  
**Database:** Compendex  
[Abstract](#) - [Detailed](#)
  
- 4. **Report: NAB 2005**  
[Wood, David](#) (European Broadcasting Union) **Source:** *EBU Technical Review*, n 303, July, 2005, 10p  
**Database:** Compendex  
[Abstract](#) - [Detailed](#)
  
- 5. **Training librarians in the production of distance learning materials: experiences of the PROLIB project**  
[Tedd, Lucy](#) (Univ of Wales); [Tetrevova, Milena](#); [Thomas, Clare](#) **Source:** *Education for Information*, v 18, n 1, Mar, 2000, p 67-76  
**Database:** Compendex  
[Abstract](#) - [Detailed](#)
  
- 1. Nicholson S, Lankes RD  
[The digital reference electronic warehouse project - Creating the infrastructure for digital reference research through a multidisciplinary knowledge base](#)  
REFERENCE & USER SERVICES QUARTERLY 46 (3): 45-59 SPR 2007  
Times Cited: 0
  
- 2. Henschke E  
[Digitising the hand-written bible: The Codex Sinaiticus, its history and modern presentation](#)  
LIBRI 57 (1): 45-51 MAR 2007

Times Cited: 0

- 3. Borges JC, Cagliari TC, Ramos CHI  
[E-Expression and variability of molecular chaperones in the sugarcane expressome](#)  
JOURNAL OF PLANT PHYSIOLOGY 164 (4): 505-513 APR 2007  
Times Cited: 0
  
- 4. Hojnacki SM, Kastner JH, Micela G, et al.  
[An X-ray spectral classification algorithm with application to young stellar clusters](#)  
ASTROPHYSICAL JOURNAL 659 (1): 585-598 Part 1 APR 10 2007  
Times Cited: 0
  
- 5. Tumlin M, Harris SR, Buchanan H, et al.  
[Collectivism vs. individualism in a wiki world: Librarians respond to Jaron Lanier's essay "Digital Maoism: The Hazards of the New Online Collectivism"](#)  
SERIALS REVIEW 33 (1): 45-53 MAR 2007  
Times Cited: 0
  
- 6. De Santo M, Percannella G, Sansone C, et al.  
[Segmentation of news videos based on audio-video information](#)  
PATTERN ANALYSIS AND APPLICATIONS 10 (2): 135-145 MAY 2007  
Times Cited: 0
  
- 7. Kwon N  
[Public library patrons' use of collaborative chat reference service: The effectiveness of question answering by question type](#)  
LIBRARY & INFORMATION SCIENCE RESEARCH 29 (1): 70-91 2007  
Times Cited: 0
  
- 8. Moskovitch R, Martins SB, Behiri E, et al.  
[A comparative evaluation of full-text, concept-based, and context-sensitive search](#)  
JOURNAL OF THE AMERICAN MEDICAL INFORMATICS ASSOCIATION 14 (2): 164-174  
MAR-APR 2007

Times Cited: 0

- 9. Crawford J, Irving C

[Information literacy: The link between secondary and tertiary education project and its wider implications](#)

JOURNAL OF LIBRARIANSHIP AND INFORMATION SCIENCE 39 (1): 17-26 MAR 2006

Times Cited: 0

- 10. Arnott G

[Digital library development: The view from Kanazawa](#)

JOURNAL OF LIBRARIANSHIP AND INFORMATION SCIENCE 39 (1): 59-59 MAR 2006

Times Cited: 0

11.) Adaptive and intelligent web based education system: Towards an integral architecture and framework

*Expert Systems with Applications, Volume 33, Issue 4, November 2007, Pages 1076-1089*

Alejandro Canales, Alejandro Peña, Rubén Peredo, Humberto Sossa and Agustín Gutiérrez

Abstract

In this paper it is presented our contribution for carrying out adaptive and intelligent Web-based Education Systems (WBES) that take into account the individual student learning requirements, by means of a holistic architecture and Framework for developing WBES. In addition, three basic modules of the proposed WBES are outlined: an Authoring tool, a Semantic Web-based Evaluation, and a Cognitive Maps-based Student Model. As well, it is stated a Service Oriented Architecture (SOA) oriented to deploy reusable, accessible, durable and interoperable services. The approach enhances the Learning Technology Standard Architecture, proposed by IEEE-LTSA (Learning Technology System Architecture) [IEEE 1484.1/D9 LTSA (2001). Draft standard for learning technology – learning technology systems architecture (LTSA). New York, USA. URL: <http://ieee.ltsc.org/wg1>], and the Sharable Content Object Reusable Model (SCORM), claimed by Advanced Distributed Learning (ADL) [Advanced Distributed Learning Initiative (2004). URL: <http://www.adlnet.org>].

## 9.3 Uporabljena programska orodja

- 1.) Orange Canvas (*complete open source*)
- 2.) Ilog Discovery (*z dovoljenjem proizvajalcev*)
- 3.) AntConc 3.1.302 (*complete open source, z dovoljenjem proizvajalca*)
- 4.) CBA 1.0 (*z dovoljenjem proizvajalcev*)
- 5.) UML Visual Paradigm Community edition (*z dovoljenjem proizvajalcev*)
- 6.) Microsoft Visio Professional 2007 (*z akademsko licenco*)
- 7.) MIDOS Thesaurus 2000 (*z dovoljenjem proizvajalcev*)

## 10 Vsebina CD-ROM-a

- 1.) Doktorsko delo na CD-ROM
- 2.) Meritve, analize idr. (*gre za zbrane podatke poizvedb in analize*)
- 3.) Modeli (*zbirka različnih slikovnih prikazov in preizkusov*)
- 4.) Prototip MRT (*kvalitativen model adaptivne digitalne knjižnice. Za ogled te prototipne izvedbe je potrebno programsko orodje UML Visual Paradigm 6.2 Community edition – obseg okoli 130 MB*). Znotraj tega paketa se nahaja tudi prototipna izvedba multidimenzionalnega rang tezavra, ki sem jo izdelal s pomočjo programskega orodja MIDOS Thesaurus 2000.

# Izjava

Izjavljam, da sem doktorsko delo izdelal samostojno pod vodstvom mentorja izred. prof. dr. Marjana Krisperja in somentorja red. prof. dr. Vladislava Rajkoviča.

Ljubljana, 21.05.2008

Karl Petrič

---

# Zahvala

Najlepše se zahvaljujem vsem, ki so pripomogli k nastanku tega dela: predvsem svojemu mentorju, izred. prof. dr. Marjanu Krisperju s Fakultete za računalništvo in informatiko v Ljubljani in somentorju red. prof. dr. Vladislavu Rajkoviču (*Fakulteta za organizacijske vede Kranj*), ki sta to raziskavo in njeno objavo spodbudila. Povrhu tega se še zahvaljujem izred. prof. Viljanu Mahniču s Fakultete za računalništvo in informatiko, red. prof. dr. Józsefu Györkösu, red. prof. dr. Samu Kralju (*Naravoslovna tehniška fakulteta v Mariboru*) in bratu, izred. prof. dr. Teodorju Petriču (*Filozofska fakulteta v Mariboru*). Ob tej priložnosti bi se še zahvalil svojim sodelavkam v specialni knjižnici na Ministrstvu za notranje zadeve (*v tem vpogledu naj še zlasti omenim sodelavko gospo Danico Hanžič, s katero sva mnogo razpravljala o tezavrih in vodjo knjižnice gospo Anteo Kursar-Trček, ki je s sodobnim sodelovalnim slogom vodenja, vedno spodbudila konstruktivne predloge*) za njihovo razumevanje in spodbude pri mojih raziskovalnih prizadevanjih. Velika zahvala gre tudi gospodu informacijskemu specialistu Marku Šercelju iz Centralne tehniške knjižnice v Ljubljani, ki je s poizvedbami po različnih podatkovnih zbirkah potrdil izvirnost teme mojega doktorskega dela, kajti izidi njegovih poizvedb so bili identični z mojimi. Prav tako se zahvaljujem vsem dejavnim spletnim obiskovalcem pilotne digitalne knjižnice, ki so obiskali te strani in izvedli zame dragocene poizvedbe, na podlagi katerih sem lahko več let zbiral podatke in jih v poznejšem obdobju obdelal. Še zlasti se zahvaljujem vsem proizvajalcem kakovostnih (*brezplačnih*) programskih orodij kot so UML Visual Paradigm Community edition 6.0/6.1/6.2, Ilog Discovery, CBA 1.0, Orange Canvas, Microsoft Visio Professional 2007, MIDOS Thesaurus 2000 in AntConc, ki so mi omogočila, da sem lahko v dobrobit tega dela zmožl izvesti dragocene in poučne analize!

Nenazadnje se zahvaljujem staršem, ki sta mi z vsestransko podporo in potrpežljivostjo sploh omogočila študij in nemoteno delo.