

UNIVERZA V LJUBLJANI
FAKULTETA ZA RAČUNALNIŠTVO IN INFORMATIKO

Janez Puntar

**Funkcionalnosti in primerjava
učinkovitosti iskalnikov platforme
SharePoint**

DIPLOMSKO DELO

VISOKOŠOLSKE STROKOVNE ŠTUDIJSKE PROGRAM PRVE
STOPNJE RAČUNALNIŠTVO IN INFORMATIKA

MENTOR: viš. pred. dr. Aljaž Zrnec

Ljubljana, 2012

Rezultati diplomskega dela so intelektualna lastnina avtorja in Fakultete za računalništvo in informatiko Univerze v Ljubljani. Za objavlanje ali izkoriščanje rezultatov diplomskega dela je potrebno pisno soglasje avtorja, Fakultete za računalništvo in informatiko ter mentorja.

Besedilo je oblikovano z urejevalnikom besedil \LaTeX .



Št. naloge: 00216/2012

Datum: 04.04.2012

Univerza v Ljubljani, Fakulteta za računalništvo in informatiko izdaja naslednjo nalogo:

Kandidat: **JANEZ PUNTAR**

Naslov: **FUNKCIONALNOSTI IN PRIMERJAVA UČINKOVITOSTI ISKALNIKOV
PLATFORME SHAREPOINT**

**THE FUNCTIONALITIES AND PERFORMANCE COMPARISON OF
SHAREPOINT SEARCH ENGINES**

Vrsta naloge: Diplomsko delo visokošolskega strokovnega študija prve stopnje

Tematika naloge:

V današnjem času se v številnih poslovnih sistemih ustvarja velika količina dokumentov, katere morajo izvajalci poslovnega procesa pregledati, uporabiti in shranjevati za kasnejšo rabo. Zaradi tega pride do razdrobljenosti in porazdeljenosti dokumentov po različnih računalnikih v poslovnem sistemu. V okviru diplomske naloge predstavite platformo Sharepoint, ki poskrbi, da si zaposleni lahko na enem mestu delijo podatke in pa iskalnik za iskanje po portalu Search Server. Realizirajte integracijo iskalnika v portal Sharepoint ter predstavite zmogljivosti in pomanjkljivosti iskalnika. Predstavite tudi odpravo prej omenjenih pomanjkljivosti z orodjem Search server express.

Mentor:

viš. pred. dr. Aljaž Zrnec



Dekan:

prof. dr. Nikolaj Zimic

IZJAVA O AVTORSTVU DIPLOMSKEGA DELA

Spodaj podpisani Janez Puntar, z vpisno številko **63090281**, sem avtor diplomskega dela z naslovom:

Funkcionalnosti in primerjava učinkovitosti iskalnikov platforme Share-Point.

S svojim podpisom zagotavljam, da:

- sem diplomsko delo izdelal samostojno pod mentorstvom viš. pred. dr. Aljaža Zrneca,
- so elektronska oblika diplomskega dela, naslov (slov., angl.), povzetek (slov., angl.) ter ključne besede (slov., angl.) identični s tiskano obliko diplomskega dela,
- soglašam z javno objavo elektronske oblike diplomskega dela v zbirki "Dela FRI".

V Ljubljani, dne 11. novembra 2012

Podpis avtorja:

Iskreno se zahvaljujem svojemu mentorju viš. pred. dr. Aljažu Zrnecu na Fakulteti za računalništvo in informatiko v Ljubljani za strokovno pomoč in nasvete pri izdelavi diplomske naloge. Zahvalil bi se tudi Zavodu Arnes, ker mi je omogočil postavitev infrastrukture, na kateri sem lahko izvajal testiranja in poizkuse.

Posebna zahvala za podporo skozi celotno študijsko pot pa gre mojim najbližjim.

Nazadnje se zahvaljujem vsem ostalim, ki so kar koli prispevali k nastanku diplomske naloge in uspešnemu zaključku študija.

Kazalo

Povzetek

Abstract

1	Uvod	1
2	Pregled platform SharePoint	3
2.1	SharePont Foundation 2010	3
2.2	SharePoint Server 2010: Intranet	4
2.3	SharePoint Server 2010: Internet/Extranet	5
2.4	Vrste omrežij glede na dostopnost	5
2.5	Zaključek	6
3	Pregled iskalnikov SharePoint	7
3.1	SharePoint Foundation 2010 search	7
3.2	Search Server 2010 Express	8
3.3	Search Server 2010	8
3.4	FAST Search Server	9
4	Arhitektura iskalnikov SharePoint	13
5	Delovanje delov iskalnika	17
5.1	Pregledovalnik (Crawler)	17
5.2	Indeksator	18
5.3	Poizvedovalnik (Query Processor)	18

5.4	Podatkovna baza	18
6	Skalabilnost sistema	21
6.1	Strojna oprema	22
6.2	Izračun velikosti podatkovnega prostora	23
7	Strukture omrežij	25
8	Uporabljena orodja	29
8.1	Microsoft Visual Studio 2010	29
8.2	Microsoft .NET Framework	30
8.3	Management studio	31
8.4	IIS	31
8.5	Microsoft SharePoint Designer 2010	31
9	Implementacija	33
9.1	Varnost	33
9.2	Pravice	35
9.3	Testni podatki	35
9.4	Organizacija SharePointa	39
9.5	BCS	39
9.6	Nastavitve pregledovalnika	40
9.7	Uporaba iskalnika	41
9.8	Ukazna vrstica	43
9.9	Posodobitev podatkovne baze	43
10	Zaključek	45

Seznam uporabljenih kratic in simbolov

CMS - Content Management System

WSS - Windows Sharepoint Service

ECM - Enterprise Content Management

AD - Active Directory

HTML - HyperText Markup Language

HTTP - HyperText Transfer Protocol

CAL - Client Access License

ECAL - Enterprise Client Access License

ACL - Access Control List

CLR - Common Language Runtime

XML - Extensible Markup Language

CSS - Cascading Style Sheets

NTLM - NT LAN Manager

OCR - Optical Character Recognition

Povzetek

Diplomska naloga opisuje primerjave, integracijo in uporabo iskalnika na portalu SharePoint. V diplomski nalogi smo najprej preučili, zakaj skupine ljudi potrebujejo stične točke, za optimalno delovanje na njihovem področju. Izkaže se, da ko imajo skupine stične točke velikokrat nastane zmeda, kako in kje so shranjeni podatki. V ta namen smo postavili svoj portal in iskalni strežnik. Z uporabo naše infrastrukture smo odpravili veliko večino problemov, ki se pojavijo v produkcijskem okolju. Težave smo razdelili v tri skupine. Zanimale so nas preformančne, funkcijske in varnostne zmožnosti. V sklopu performančnih zmožnosti smo si ogledali različne strukture sistemov ter priporočila, kdaj je treba izbrati posamezno. Ker imamo na izbiro kar pet različnih iskalnikov, se moramo odločiti za takega, ki bo zadostoval našim potrebam, a bo še vedno finančno ugoden. Vse to smo zajeli v sklopu funkcionalnosti iskalnikov. Na portalu smo se spoznali tudi z varnostno politiko in razširjanjem portala s svojimi programskimi deli.

Abstract

The thesis describes the comparison, integration and use the search engine on SharePoint portal. In this thesis, we first examined why groups of people need the contact points for optimal performance in their field. In cases when we have a lot of users and data, sometimes individuals don't know where the files are stored and they create another one. This leads to confusion with a lot of similar files. To this end, we set up a portal and search server. Using our infrastructure, we eliminated the vast majority of problems that occur in a production environment. Problems were divided into three groups. We were interested in performance, functional and security capabilities. As part of performance capabilities we visited various topologies systems and recommendations when to choose each. Since we have five options to choose from different search engines, we have to decide on one that will be sufficient for our needs but will still be financially beneficial. All this was captured as part of the functionality of search engines. On SharePoint portal we also met with security policies and development our own webparts.

Poglavje 1

Uvod

Microsoft je prvo spletno platformo SharePoint izdal leta 2001. Namenjena je bila shranjevanju in urejanju dokumentov. Takšni portali spadajo pod skupino CMS (content management system). Sedaj je na voljo že četrta različica, SharePoint 2010. Vsebuje številne novosti, katere lahko skupine in podjetja izkoristijo za hitrejše in še bolj organizirano delovanje.

V Microsoftovemu oddelku za prodajo SharePoint opisujejo kot »Business Collaboration Platform for the Enterprise and the Web« in utemeljujejo, da:

- povezuje in pomaga skupinam ljudi,
- znižuje stroške s poenoteno infrastrukturo,
- omogoča hitro in prožno prilagajanje poslovnim potrebam.

SharePoint je obsežen in tudi zelo prilagodljiv portal, tako da se lahko sami odločimo, katere komponente (Slika 1.1) bomo uvedli v njemu. Z večanjem integracije poslovnega procesa ciljnega podjetja se izboljšuje tudi uporabnost portala [2].

Portal nam omogoča integracijo poslovne logike, upravljanja s podatki, povezovanja skupin ljudi in iskanja po portalu. Iskalnik je zelo pomemben del platforme. Uporabniki portalov pričakujejo, da bodo hitro našli zelene shranjene informacije. Rezultati raziskav IDC-ja (International Data Corporation) iz leta 2007 kažejo, da zaposleni v povprečju porabi 9,5 ur službenega



Slika 1.1: Komponente platforme SharePoint [19].

časa na teden za iskanje že obstoječih informacij. Še bolj skrb vzbujajoče dejstvo pa je, da zaposleni v povprečju porabi 6 ur na teden za kreiranje dokumentov, ki že obstajajo, a jih ne more najti. Uspešnost iskanja informacij pa je le štirideset procentna. Ob tako visokem številu neproduktivnih ur zaposlenih podjetje iz dneva v dan izgublja dobiček. Iz navedenih podatkov lahko sklepamo, da zaposleni potrebujejo orodje, ki jim omogoča, da na enem mestu najdejo vse potrebne informacije.

Glavni namen diplomske naloge je raziskati različne vrste iskalnikov. Izbira prave različice iskalnika je pomembna tudi s finančnega vidika. Odločiti se moramo za takšno strukturo, ki bo ustrezala našim potrebam, a bo še vedno cenovno ugodna. Raziskave v diplomski nalogi nam bodo prišle v pomoč tudi pri predstavljanju produktov strankam in kasnejši zaposlitvi.

Poglavje 2

Pregled platform SharePoint

Microsoft ima pripravljenih nekaj različic [1] portala SharePoint, nekatere že vsebujejo integriran iskalnik, v ostalih pa lahko iskalnik integriramo sami. Iskalnik lahko kupimo ali pa uporabljamo brezplačnega. V tem poglavju bomo naredili pregled vseh različic platform in katere funkcionalnosti vsebuje posamezna različica.

2.1 SharePont Foundation 2010

Obstajata dve verziji platforme SharePoint. Prva, z imenom SharePoint Foundation 2010, je brezplačna. Pogoji za uporabo je, da imamo licenčen operacijski sistem Windows Server 2008. To verzijo smo uporabili pri implementaciji iskalnika. Več o implementaciji bomo predstavili v nadaljevanju.

Z brezplačno verzijo pridobimo vse glavne funkcionalnosti. Ena večjih sprememb v brezplačni verziji je vgradnja BCS-ja (Business Connectivity Services). Omogoča nam interakcijo z ostalimi Microsoftovimi in drugimi produkti. Če želimo urejati tekstovne datoteke, jih lahko kar preko orodja Office Word. Urejevalnik besedil Word se v tem primeru poveže na strežnik in pridobi zahtevano datoteko. Če nam je SharePoint všeč, ga lahko preprosto nadgradimo na verzijo, imenovano SharePoint Server 2010.

Ob namestitvi brezplačne verzije dobimo tudi vgrajen iskalnik. Iskalnik

ima zelo omejene funkcionalnosti. Išče lahko samo po enem naboru strani na portalu.

2.2 SharePoint Server 2010: Intranet

SharePoint Server 2010 je naslednik brezplačne verzije. Za to verzijo se odločajo podjetja, ki želijo bolj napredne storitve njihovega internega portala. Izbiramo lahko med dvema različicama: Standard in Enterprise. Razliki med tema dvema verzijama sta cena licenciranja in nabor funkcionalnosti. Ob namestitvi je integriran tudi zelo izpopolnjen iskalnik.

Standard

Najpomembnejše funkcionalnosti verzije Standard so:

- Enterprise Content Management (ECM): Funkcionalnost poskrbi, da so podatki ločeni od stila strani. Tako lahko oblikovalci in razvijalci portala neodvisno opravljajo vsak svoje delo.
- Centralizacija SharePoint nastavitvev vseh strežnikov v omrežju.
- Izpopolnjen iskalnik. Uporabnikom omogoča označevanje vsebine in ocenjevanje, kar izboljša delovanje iskalnika. Iskalnik lahko indeksira približno 100 milijonov zapisov.
- Uporabniki lahko portal uporabljajo kot osebni računalnik. Na portalu lahko ustvarjajo mape in dokumente. Vsi podatki pa se shranjujejo na strežnik s pomočjo tehnologije Active Directory.
- Portal in iskalnik lahko uporabljamo tudi kot socialno omrežje.

Enterprise

- Izboljšana storitev InfoPath. Omogoča nalaganje in pošiljanje form preko strežnika, kar sprosti resurse pri odjemalcu.

- Izkorišča Excel in PowerPoint datoteke kot vir podatkov za druge aplikacije.
- PerformancePoint: Omogoča centralni pregled delovanja omrežja in storitev.
- Omogoča prenos Accessove podatkovne baze v okolje SharePoint, ker je zagotovljeno bolj podprto in centralno nadzorovano delovanje.
- Omogoča prenos in obdelavo podatkov Visio na portal SharePoint.
- Integriran je tudi iskalnik »FAST Search«. Ponuja še boljše nastavitve pri iskanju po portalu.

2.3 SharePoint Server 2010: Internet/Extranet

Različica za extranet in internet ima popolnoma enake funkcionalnosti kot za intranet. Razlika je samo v licenciranju, ker ne moremo plačati licenc za vse uporabnike, tako kot to naredimo za interne sisteme s pogodbo CAL (Client Access License).

2.4 Vrste omrežij glede na dostopnost

- Intranet [9] je zaprto omrežje, sestavljeno iz računalnikov posamezne organizacije. Ti računalniki so povezani med seboj in imajo načeloma samo eno izhodno točko do svetovnega spleta. Ves promet, ki ga organizacija proizvede, mora prečkati požarni zid. S tako arhitekturo omrežja je organizacija zelo zavarovana, ker nihče s svetovnega spleta ne more dostopati do podatkov.
- Internet je najbolj razširjena oblika omrežja. Običajno se uporablja za predstavitve organizacij in informacij, namenjenih širšim množicam.

- Ekstranet je mešanica interneta in intraneta. Podatki v omrežju so pomembni za delovanje ustanove in jih ne želijo deliti z vsemi. Do njih lahko dostopajo samo pooblašcene osebe, organizacije in podjetja. Primer takšne organizacije bi bilo podjetje, ki izdeluje avtomobile. Podjetje dovoli dobaviteljem, da vnašajo in pregledujejo podatke v zvezi z dobavo na njihovem omrežju.

2.5 Zaključek

V tem poglavju smo naredili pregled različic portala SharePoint. Spoznali smo se tudi s politiko licenciranja portala [5]. Pred začetkom implementiranja iskalnikov in celotnega portala je zelo zaželeno temeljito posvetovanje s strokovnjaki licenciranja SharePointovih izdelkov.

Poglavje 3

Pregled iskalnikov SharePoint

V tem poglavju bomo podrobno opisali funkcionalnosti, razširljivosti in zmogljivosti posameznega iskalnika. Zanimalo nas bo, kdaj se odločimo za posamezno različico in kakšne funkcionalnosti ponuja.

3.1 SharePoint Foundation 2010 search

Različico iskalnika dobimo že vgrajeno ob namestitvi portala SharePoint Foundation [6]. To je najbolj okrnjena verzija iskalnika. Iskanje nam omogoča samo na spletnih straneh, ki jih naredimo ali uvozimo na portal. Z iskalnikom ne moremo iskati po virih, kot so datoteke, shranjene na disku, podatki, vneseni preko WebPartov, in podatki od drugih udeležencev portala. Ker iskalnik nima omogočenih veliko zmogljivosti, tudi nastavitve niso potrebne. To bi lahko spadalo pod pluse, če ne želimo imeti opravka z optimizacijo iskalnika, drugače pa je to velika slabost. Kljub temu da nam iskalnik ne omogoča veliko zmogljivosti, ga lahko še vedno uporabimo v primeru, ko podatke shranjujemo v obliki spletnih strani, ko nimamo datotek, shranjenih izven portala, in v primeru ko ne želimo investirati v dodatno infrastrukturo, ki bi pospešila in izboljšala točnost iskalnih rezultatov. Vsak iskalni strežnik nam omogoča indeksiranje 10 milijonov zapisov.

3.2 Search Server 2010 Express

Ta verzija [2] je namenjena zahtevnejšemu okolju. Iskalnik ima skoraj enake možnosti iskanja in priprave indeksov kot SharePoint Server 2010. Dejstvo, da ima zelo podobne funkcionalnosti kot SharePoint Server 2010, pripomore k temu, da uporabniki-raziskovalci za testne in produkcijske namene najprej namestijo Express verzijo. Omogoča nam iskanje po datotekah izven platforme, spletnih straneh, po več zbirkah spletnih strani, po Lotus Notesu in po podatkih, ki smo jih vnesli preko WebPartov (Business Data Catalog).

Slabost verzije Express je, da lahko implementiramo samo eno instanco iskalnika. Problem pri takšni strukturi omrežja nastane ob izpadu iskalnega strežnika ali pa pri preobremenitvah. Iz sledečega je razvidno, da ima Microsoft dobro izdelano tržno taktiko. Če želimo imeti redundanten in zmogljiv iskalni strežnik z veliko večino funkcionalnosti, bomo za njega morali plačati.

Poleg omejitve števila strežnikov nas iskalni strežnik omejuje tudi v številu zapisov v podatkovni bazi. Če za podatkovno bazo uporabimo SQL Server Express, lahko iskalni strežnik indeksira 300.000 zapisov. V primeru, da se odločimo za nakup plačljive verzije strežnika SQL, pa število zapisov naraste na 10 milijonov.

3.3 Search Server 2010

Search Server 2010 je nadgradnja verzije Express. V njej ni veliko novih funkcionalnosti. Prednosti v tej različici se navezujejo predvsem na performančne in skalabilne lastnosti. Zelo primerna je za velika poslovna okolja. S to različico lahko poljubno prirejamo strukturo omrežja, dokler ne bo zadovala našim potrebam. Indeksiramo lahko približno 100 milijonov zapisov. Več o vrstah strukturah omrežij si bomo pogledali v nadaljevanju.

3.4 FAST Search Server

Strežnik vsebuje vse iskalne funkcionalnosti platforme SharePoint Server 2010. Dodatno vrednost pridobi z zelo visoko prilagodljivostjo, skalabilnostjo in izboljšanimi metodami pri pripravljanju in pregledovanju vsebine, ki jo uporablja iskalnik. Ko v poslovnem okolju presežemo 500 milijonov zapisov, nam Microsoft priporoča FAST Search Server. Iskalni strežnik je tako optimiziran, da ga moramo namestiti na ločenem strežniku. Z iskalnikom lahko indeksiramo približno en bilijon zapisov. Za to verzijo iskalnika se običajno odločajo zelo velika podjetja, ki morajo imeti z Microsoftom sklenjene »Enterprise« licence, imenovane ECAL.

V tabeli 3.1 smo združili funkcionalnosti vseh iskalnikov, jih opisali, ter označili v kateri različici iskalnika jih je mogoče implementirati.

Tabela 3.1: Prikaz in opis funkcionalnosti različnih verzij iskalnikov [7].

	SP Foun- dation 2010	Search Server 2010 Express	Search Server 2010	SP Server 2010	FAST Search Server 2010
Osnovno iskanje.	✓	✓	✓	✓	✓
Visual Best Bets. Besede in besedne zveze, katere administrator določi glede na statistiko preteklega iskanja.		Omej.	Omej.	Omej.	✓
Obseg iskanja. Uporabnik lahko izbere obseg iskanja, da dobi še bolj filtrirane rezultate.		✓	✓	✓	✓
Izboljšave iskalnika, ki temeljijo na kontekstu uporabnika.					✓

Vključuje besede in besedne zveze, ki ji izbere administrator, ponuja najbolj iskane dokumente glede na iskalni niz, išče dokumente glede na skupino, v kateri je uporabnik.

Nastavitve delovanja iskalnika.		✓	✓	✓	✓
Administrator lahko določa, kateri zapisi se bodo indeksirali in kako bodo prikazani iskalni rezultati.					
Meta podatki posameznih zapisov.		Omej.	Omej.	Omej.	✓
Iskalnik pri indeksiranju preišče in shrani tudi ostale podatke o zapisih, ne samo vsebino (avtor, naslov, datum nastanka). Iskalni rezultati so posledično bolj točni.					
Query federation.		✓	✓	✓	✓
Združuje rezultate iskanja z več iskalnih strežnikov.					
Predlogi pri vpisovanju iskalnega niza.		✓	✓	✓	✓
Similar results.					✓
Generira nove rezultate na podlagi izbranih zapisov prejšnjih poizvedb.					
Prioritiziranje rezultatov.					✓
Urejanje rezultatov glede na nastavljene lastnosti ali s pomočjo FQL formule (FAST Query Language).					
Povezovanje med iskalnimi pojmi in vsebino.		Omej.	Omej.	Omej.	✓
Določanje posameznih besednih zvez relevantnim skupinam dokumentov, spletnih strani in ostalih zapisov. S to funkcijo lahko tudi znižujemo rang posameznim manj pomembnim dokumentom.					
Shallow results refinement.		✓	✓	✓	✓

Izboljšanje rezultatov na podlagi meta podatkov najboljših rezultatov.					
Deep results refinement.					✓
Izboljšanje rezultatov na podlagi vseh meta podatkov.					
Predogled datotek.					✓
Hiter predogled datotek v formi z rezultati iskanja (Word, PowerPoint in Excel).					
Povezljivost z Windowsom 7.		✓	✓	✓	✓
Uporabniki lahko datoteke in ostale naprednejše zapise iščejo kar preko iskalnika v Raziskovalcu.					
Iskanje ljudi in uporabnikov portala.				✓	✓
Socialno omrežje.				✓	✓
Točnost iskalnika je izboljšana s pregledovanjem povezav med uporabniki, povezav posamezne vsebine in uporabniki ter glede na to, kako uporabniki označujejo posamezne vsebine. Zaposleni v računovodstvu uporabljajo različne vsebine kot v proizvodnji.					
Taxonomy integration.				✓	✓
Storitev uporablja predefinirane oznake, ki jih lahko pripnemo posamezni vsebini na portalu. S tem izboljšamo natančnost rezultatov (oznake, ocene).					
Porazdeljen sistem (Multi-tenant hosting).				✓	
Podatki vsake organizacije (tenant) so lahko porazdeljeni po celotnem sistemu. Storitev je primerna, če želimo gostiti več podjetij/ustanov. Podjetja/ustanove bodo delovala na isti infrastrukturi, ne bodo mogla dostopati do podatkov ostalih strank.					

Indeksiranje kompleksnejših spletnih vsebin. (Flash) - Rich Web indexing support.					✓
--	--	--	--	--	---

Tabela 3.1: Prikaz in opis funkcionalnosti različnih verzij
iskalnikov [7].

Poglavje 4

Arhitektura iskalnikov SharePoint

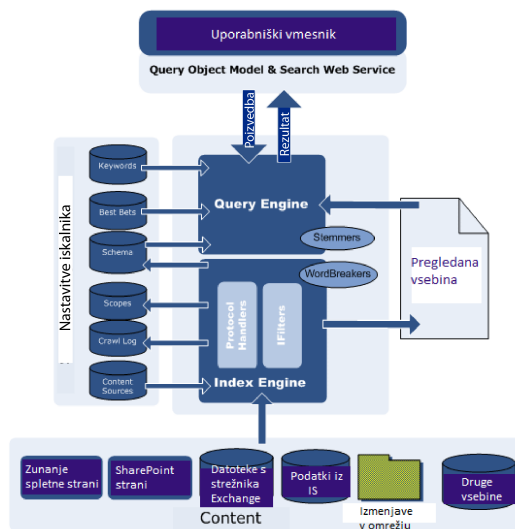
V tem poglavju bomo preučili arhitekturo [8] iskalnikov SharePoint (Slika 4.1). Poznavanje arhitekture pripomore k kasnejšemu reševanju problemov, ki se lahko pojavijo ob samem delovanju. Vse različice iskalnikov so sestavljene iz štirih glavnih komponent. Vsebujejo pregledovalnik (crawler), indeksator (indexer), razčlenjevalnik poizvedb (query processor) in podatke. Zanimivost je, da vse različice iskalnikov uporabljajo enako arhitekturo. To nam omogoča preprosto razširjanje skalabilnosti iskalnikov. Vse nastavitve, tudi v celotnem omrežju SharePoint, lahko nastavimo preko spletnega vmesnika. Razlike, ki se pojavijo med vrstami iskalnikov, so programske. Vsaka varianta drugače komunicira s komponentami iskalnika.

Query Engine

Skrbi za pravilno interpretacijo iskalnega niza in vračanje rezultatov. Delovanje poizvedb lahko omejimo na določeno področje vsebine.

Content Index

Hrani podatke o vsebini. V tej datoteki so shranjeni vsi podatki o vseh besedah. Najpomembnejše je, v kateri datoteki je shranjena in kje je ta datoteka. Za razvrščanje poskrbi indeksator.



Slika 4.1: Arhitektura iskalnika SharePoint [8].

Index Engine

Uporablja se za pregledovanje celotne vsebine dokumentov. Pregledane dokumente razvrsti v datoteko, imenovano indeks (Content Index), ki se uporablja za hitrejša iskanje vsebine. Indeksator (Index Engine) sestavljata dve komponenti. IFilter in Protocol Handlers. IFilter je zadolžen za prebiranje datotek v njihovem primarnem formatu. Po potrebi lahko uvozimo dodatne pregledovalnike. Rokovalnik protokolov (Protocol Handler) pa je zadolžen za odpiranje datotek po njihovem primarnem protokolu.

- Področja (Scopes).
Prednastavljena področja iskanja. Uporabljata jih lahko uporabnik iskalnika in sam iskalnik.
- Dnevnik (Crawl Log).
Zapisuje dnevnik poročil uspešnosti pregledovanja vsebine.
- Vir podatkov (Content Sources). Vsebuje poti do podatkov.

Vsebina (Content)

Vsebino lahko na portal SharePoint dodamo iz mnogo različnih virov:

- Zunanje spletne strani.
- SharePoint strani.
- Podatki s strežnika Exchange.
- Podatki, pridobljeni s poslovno logiko podjetja/ustanove.

Poglavje 5

Delovanje delov iskalnika

V poglavju bomo opisali povezavo in delovanje med štirimi najbolj pomembnimi komponentami iskalnika. Te komponente so ključnega pomena pri predpripravi podatkov za iskanje, kot tudi v času posamezne poizvedbe.

5.1 Pregledovalnik (Crawler)

Pregledovalnik je zelo pomemben del sestave SharePointa. Zadolžen je za zbiranje podatkov iz različnih virov vsebine. Zbrano vsebino pregleda in združi v podatkovno bazo, katero kasneje uporablja indeksator. Način shranjevanja podatkov v tej fazi ni pomemben. Pomembno je le, da so pregledani vsi podatki. Da pregledovalnik lahko preišče vse vrste podatkov, mora uporabljati rokovalnik protokolov. Pregledovalniku damo na voljo lahko tudi podatke o osebah. Črpamo jih iz podatkovne strukture, imenovane Active Directory – AC. Active Directory temelji na protokolu LDAP, kar pripomore k varnosti podatkov. Administrator portala lahko določi, kateri podatki bodo vidni posamezni osebi in pregledovalniku.

5.2 Indeksator

Podatki so pripravljani na portalu SharePoint, a niso pregledani in urejeni. Za primer vzemimo datoteko s končnico doc. Pregledovalnik (Crawler) zazna datoteko, jo odpre in shrani vsebino na portal. Pregledovalnik ima možnost pregleda skoraj petdeset vrst datotek. Lahko dodajamo tudi nove. Vsebina datoteke je sedaj na portalu, a jo iskalnik ne razume. V tej točki nastopi indeksator (Index Engine), ki podatke predela v logične strukture, razumljive poizvedovalniku.

5.3 Poizvedovalnik (Query Processor)

Poizvedovalnik je sestavljen iz več delov. Eden izmed njih je tudi vnosno polje, v katerega vpišemo iskalni niz. Iskalni niz mora biti najprej programsko obdelan. Programska obdelava vključuje razbijanje zloženk, besed, sestavljenih iz več korenov, na posamezne korene besed, sestavljanje logičnih operacij vnesenega niza, postavitve določenih znakov ob posamezni besedi/stavku, v primeru, če ne vemo, kako se beseda/stavek konča ali začne. Tako kot k vsaki strojni obdelavi jezika spada korenjenje, uvedli so ga tudi v iskalnikih SharePoint. Poizvedovalnik je tudi element celotnega iskalnika, pri katerem je največ razlik med različnimi verzijami. Za enak iskalni niz lahko dobimo različne rezultate. Rezultati so odvisni od povezovanja indeksne datoteke z iskalnim nizom in prioritiziranja rezultatov.

5.4 Podatkovna baza

Zelo dobra lastnost celotnega portala SharePoint je ta, da shranjuje vse podatke v podatkovno bazo [11]. S pojmom podatki tokrat opisujemo podatke, vnesene s strani uporabnikov, nastavitve portala ter bazo podatkov, ki hrani nastavitve iskalnikov in poti do datotek (Access Control List – ACL). Lastnost je dobra, ker so tako vsi podatki lažje prenosljivi po več strežnikih. Tudi kreiranje varnostne kopije podatkov je preprostejše. To lahko storimo v

centralnem administrativnem sistemu (CA – Central Administration). Nadzorni sistem nam omogoča celo, da izbiramo med strežniki v našem omrežju. Strežnike s podatkovnimi bazami lahko z namenom redundančnosti tudi zrcalimo in združujemo v gruče.

Poglavje 6

Skalabilnost sistema

V primerjavah različnih verzij portalov in iskalnikov SharePoint smo omenili njihove zmožnosti razširitev. V tem poglavju pa bomo podrobneje preučili, kakšno infrastrukturo in arhitekturo je priporočljivo implementirati za določeno količino zapisov. Ravno zaradi velikega števila možnih kombinacij različnih komponent SharePoint iskalnika je del načrtovanja zelo pomemben. Pri implementaciji iskalnika moramo biti pozorni na več faktorjev. Eden pomembnejših je hitrost delovanja. Priporočljivo je, da se uporabniku iskalni rezultati prikažejo najkasneje v eni sekundi. Pozorni moramo biti tudi na to, kakšno količino podatkov bo hranila naša stranka. To je pomembno iz dveh razlogov. Prvi je ta, da predvidimo, koliko časa se bo izvajalo pregledovanje vsebine (crawl). Drugi pa je, da bomo kupili zadostno količino prostora za hranjenje podatkov na trdih diskih. Problem, ki se lahko pojavi, če imamo prepočasno pregledovanje vsebine (crawl), je, da ko neka oseba na portal shrani novo vsebino, ta zaradi počasnosti pregledovanja še ni dostopna iskalniku, posledično tudi drugim uporabnikom portala. Težavo odpravimo z več in hitrejšimi strežniki. Vsak pregledovalni strežnik lahko tudi nastavimo na določeno vsebino ter določimo pogostost pregledovanja. Pozorni pa moramo biti tudi na varnost sistema. V primeru, da ima organizacija na portalu shranjene pomembne podatke o njihovem delovanju, moramo uporabiti varne povezave oziroma sistem postaviti interno.

Zgradbo SharePoint portala lahko zaradi njegove zgradbe razdelimo na štiri sklope. V poglavju Delovanje in arhitektura smo omenili, da so za delovanje potrebne storitve indeksiranja, pregledovanja podatkov (crawl), obdelava iskalnih nizov in shramba podatkov. Te operacije lahko razdelimo na več namenskih virtualnih ali fizičnih strežnikov. Storitve lahko ločimo na:

- spletni strežnik,
- poizvedovalni strežnik,
- pregledovalni strežnik,
- podatkovni strežnik.

6.1 Strojna oprema

Za optimalno delovanje [13] iskalnih strežnikov Microsoft priporoča da uporabimo naslednje konfiguracije:

- 4-jedrne centralno procesne enote,
- 8 GB RAM-a,
- 80 GB systemskega trdega diska.

Pri podatkovnih strežnikih lahko zaradi optimalne konfiguracije upoštevamo tudi količino podatkov. Če imamo sistem z manj kot 10 milijoni zapisov, uporabimo:

- 4-jedrne centralno procesne enote,
- 8 GB RAM-a,
- 80 GB systemskega trdega diska,
- podatkovni prostor glede na potrebe.

V primeru, ko predvidimo, da bo sistem presegal 10 milijonov zapisov, posameznim strežnikom dodelimo 16 GB pomnilnika.

6.2 Izračun velikosti podatkovnega prostora

Microsoft nam je pripravil informativen izračun [14], koliko podatkovnega prostora moramo predvideti.

Upoštevati moramo naslednje dejavnike: količino dokumentov (D), povprečno velikost dokumenta (S), število zapisov v list tabelah (L) in število verzij posameznega dokumenta brez trenutne verzije (V). 10 KB je konstanta, ki vsebuje metadata podatke.

$$\begin{aligned} \text{velikostpodatkovnebaze} &= ((D \times V) \times S) + (10KB \times (L + (V \times D))) \\ ((200.000 \times 2) \times 250) &+ ((10KB \times (600.000 + (200.000 \times 2))) = 104,9GB \end{aligned}$$

K izračunani količini podatkov moramo dodati tudi velikost administratorske podatkovne baze, baze pregledovalnika (crawl) in lastnosti. Administratorska podatkovna baza služi shranjevanju vseh nastavitev iskalnika. Velikost te baze načeloma ne bi smela prerasti 1 GB, a ji zaradi morebitnih nevspečnosti namenimo 10 GB prostora. Velikost podatkovne baze pregledovalnika pomnožimo s konstanto 0,046 in dobimo 4,83 GB. Določiti moramo tudi velikost baze z lastnostmi (metadata) o pregledanih podatkih. To določimo tako, da velikost baze pomnožimo s 0,015 in dobimo 1,575 GB.

Skupni podatkovni prostor (samo za podatke) bi v našem primeru znašal 121,4 GB.

Poglavje 7

Strukture omrežij

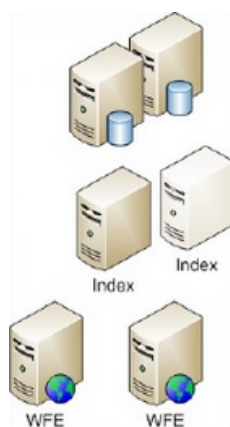
V podjetju Microsoft so za nas pripravili tudi različne strukture omrežnih sistemov [3] in rezultate [15] testiranja delovanja. V tem poglavju si bomo ogledali priporočila, kako kompleksne morajo biti strukture za določeno število datotek [2].

Če imamo manj kot en milijon dokumentov, lahko sistem postavimo samo na enem strežniku. V primeru, da že imamo podatkovni strežnik, lahko podatkovno bazo namestimo na podatkovni strežnik (Slika 7.1). S tem razbremenimo aplikacijski strežnik in izboljšamo strukturo sistema za nadaljnje nadgradnje.

V primeru enega do deset milijonov dokumentov obvezno potrebujemo



Slika 7.1: Struktura omrežja z enim strežnikom SharePoint, ki vsebuje vse vloge [?].



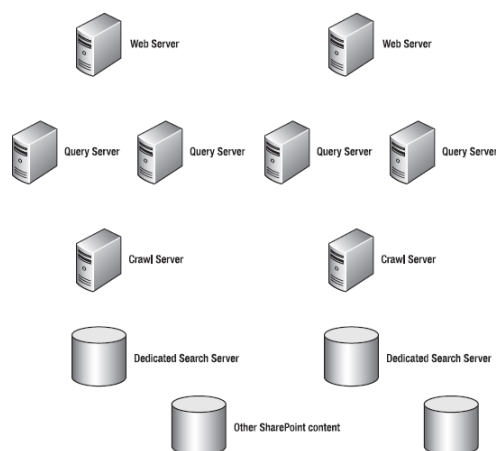
Slika 7.2: Prikaz manjše strežniške farme [16].

vsaj tri strežnike. Na enem gostimo spletni in poizvedovalni strežnik, na ostalih dveh pa imamo nameščeno podatkovno bazo in pregledovalnik (crawler) datotek. Tako strukturo lahko uporabimo, če hitrost indeksiranja ni kritična za naše potrebe. V primeru, da je hitrost indeksiranja pomembna, moramo upoštevati nasvete strukture sistema za od 10 do 20 milijonov dokumentov.

Če predvidimo, da bo naš sistem vseboval 10 do 20 milijonov dokumentov, moramo imeti štiri strežnike. Dva strežnika, ki opravljata vloge spletnega in poizvedovalnega strežnika, dva pregledovalna strežnika in strežnik s podatkovnimi bazami (Slika 7.2).

Za več kot 20 milijonov datotek je priporočljivo, da imamo namensko omrežje (strežniško farmo). Pri tem se moramo odločiti, ali bomo uporabniški vmesnik iskalnika gostili v farmi ali ne. Odločitev je pomembna, ker če se odločimo, da želimo imeti tudi uporabniški vmesnik, to pomeni, da potrebujemo tudi dva spletna strežnika. V vsakem primeru pa mora strežniška farma vsebovati štiri poizvedovalne strežnike, dva strežnika za indeksiranje in strežnike s podatkovnimi bazami (Slika 7.3). Priporočljivo je, da so podatkovne baze združene v grupo (cluster) in da imajo pomožne strežnike v primeru nenormalnega delovanja ali okvare (failover).

Naslednja stopnja se začne pri 40 do 100 milijonov dokumentov. Pri tolikšnem številu dokumentov potrebujemo strežniško farmo brez spletnih



Slika 7.3: Srednje velika strežniška farma, v kateri lahko učinkovito uporabljamo od 20 do 40 milijonov dokumentov [2].

strežnikov. Spletne strežnike moramo imeti ločene v svojem omrežju. V iskalno omrežje pošiljamo izključno zahteve za iskanje datotek in zapisov na portalu. Priporočljivo je imeti šest do deset proizvodnih strežnikov, štiri strežnike, namenjene indeksiranju, in gručo strežnikov s podatkovnimi bazami. Shema omrežja je podobna srednje veliki strežniški farmi (Slika 7.3). Razlika je, da smo ji odstranili vse spletne strežnike.

Slike v tem poglavju predstavljajo sheme omrežja za določeno število dokumentov. Strežniki z enakimi vlogami (vodoravno) so povezani skupaj in delujejo kot eden. Omrežne povezave moramo imeti tudi med strežniki z različnimi vlogami. Če smo uporabnik strukture (Slika 7.3), se vse naše proizvodne zahteve pošljejo na spletni strežnik. Zahtevo HTTP zajame strežnik z manj zahtev in jo pošlje naprej proizvodnemu strežniku. Princip delovanja je enak za vse stopnje od prejetja poizvedbe do prikazovanja rezultata.

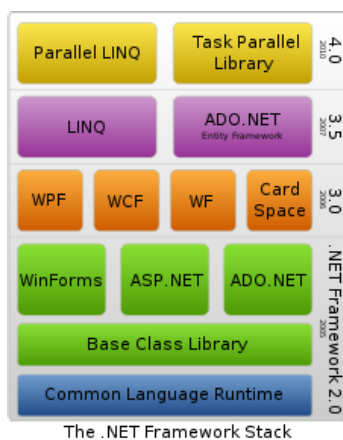
Poglavje 8

Uporabljena orodja

Pri razvoju projektov moramo biti pozorni tudi na to, kakšno podporo in razvojna orodja so nam na voljo, če želimo portal prilagoditi svojim potrebam. Microsoft ima za izdelavo in urejanje portala SharePoint zelo dobre pripomočke. V tem poglavju vam bodo predstavljena orodja, ki smo jih potrebovali za implementacijo portala.

8.1 Microsoft Visual Studio 2010

Microsoft Visual Studio [4] je integrirano razvojno orodje, ki nam omogoča razvijanje aplikacij v različnih programskih jezikih. Posamezne dele naše aplikacije smo razvijali v programskem jeziku C sharp. Visual Studio nam nudi podporo še za mnogo več programskih jezikov. Med najbolj podprte jezike spadajo še VB, C++, C, HTML, XML, JavaScript in CSS. Visual Studio nam omogoča razvoj tudi v drugih programskih jezikih, a si jih moramo namestiti sami. Zelo priročna funkcionalnost urejevalnika je IntelliSense. To je sistem, ki nam omogoča hitrejše pisanje kode. Sistem sprotno shranjuje vse naše objekte, njihove lastnosti in funkcije. Pri kasnejši uporabi pa nam avtomatsko ponudi vse shranjene možnosti iz naše programske kode in vključenih programskih knjižnic.



Slika 8.1: Prikaz verzij in strukture .NET Frameworka [10].

8.2 Microsoft .NET Framework

Za izdelavo vseh programskih delov naše aplikacije moramo imeti nameščen tudi .NET Framework [10]. Ta vsebuje veliko različnih programskih knjižnic, ki razvijalcem poenostavijo delo. Napisana programska koda deluje optimizirano v sodelovanju s strojno opremo. Microsoft funkcionalnosti Frameworku dodaja s posameznimi verzijami, tako da imamo razvijalci vedno najnovejše programske module. Osnovni razredi knjižnice (Base Class Library) nam omogočajo izgradnjo uporabniških vmesnikov, dostopa do podatkov, izdelavo kriptografije, izdelavo spletnih aplikacij, računanje numeričnih algoritmov in komunikacijo preko spleta. Poleg programskih knjižnic .NET vsebuje tudi CLR (Common Language Runtime). CLR nam omogoča, da se aplikacija zažene v navideznem okolju. Ta lastnost je dobra, ker razvijalcem aplikacij ni treba skrbeti, kakšna je strojna oprema, na kateri se bo izvajala aplikacija. Obenem pa CLR poskrbi tudi za varnost, upravljanje spomina in lovljenje izjem. Slika 8.1 nadzorno predstavlja arhitekturo in verzije sistema .NET.

8.3 Management studio

Implementirali smo tudi storitev BCS (Business Connectivity Services), ki za delovanje potrebuje podatkovno bazo. SQL Server Management Studio je idealna rešitev, če želimo s podatkovno bazo upravljati preko uporabniškega vmesnika. Omogoča nam manipuliranje s shemo podatkovne baza ter urejanje in izpisovanje podatkov.

8.4 IIS

Internet Information Services je Microsoftov spletni strežnik. Najpogosteje je uporabljen za predstavitev spletnih strani, narejenih s tehnologijo asp.net. Tudi v našem primeru smo ga uporabili s tem namenom. V spletni strežnik lahko poljubno vgradimo tudi druge spletne strežnike, če bi želeli gostiti spletne strani, narejene s tehnologijo PHP. Ob namestitvi storitve SharePoint smo poleg portala namestili tudi spletno stran centralne konzole in spletne storitve [12], ki so nepogrešljive za pravilno delovanje samega portala.

8.5 Microsoft SharePoint Designer 2010

Orodje SharePoint Designer je specialno namensko orodje za urejanje vsebine in strukture portala. S tem orodjem smo si poenostavili in pospešili delo. Omogoča nam izgradnjo celotne poslovne logike določenega podjetja, brez pisanja programske kode. To nam dopušča zelo veliko dinamičnosti in hitrega prilagajanja potrebam strank. Če bi združili skupaj portal SharePoint in SharePoint Designer, lahko storitev v določenih aspektih primerjamo tudi z Microsoftovim izdelkom Dynamics. V temu primeru še dodatno razširimo uporabo portala.

Omogoča nam tudi vizualno urejanje. Za videz spletne strani uporabljamo dve spletni tehnologiji. Prva je imenovana CSS (Cascading Style Sheets). To sedaj uporabljajo že vse spletne strani. Na običajnih spletnih straneh označimo posamezne HTML elemente in jim dodelimo razred in/ali

id. V CSS datoteki pa uporabljamo selektorje za te elemente in jim dodeljemo stil. Pri vizualnem urejanju spletnih strani portala SharePoint pa smo uporabili jezik XSL (Extensible Stylesheet Language) in XSLT (Extensible Stylesheet Language Transformation). Uporaba te spletne tehnologije nam je olajšala delo, ker lahko programiramo videz in postavitev elementov brez spreminjanja programske logike. Takšen princip načrtovanja spletnih rešitev se uporablja tudi, ko sta ekipi programerjev in oblikovalcev ločeni.

Primer uporabe označevalnega jezika v okolju SharePoint (Primer 8.1). Uporabili smo vrednost polja UnitInStock in jo primerjali, če je manjša ali enaka 30. Ob pozitivnem rezultatu se pisava tega polja spremeni v rdečkasto barvo.

```
1 <xsl:if test="$thisNode/@UnitsInStock.<=30"  
  ddwrt:cf_explicit="1">color: 71B84F;</xsl:if>
```

Listing 8.1: Uporaba označevalnega jezika v SharePointu.

Poglavje 9

Implementacija

V tem poglavju bomo realizirali namen diplomske naloge. Ustvarili bomo portal na katerem bodo zaposleni lahko opravljali svoje delo. Upoštevali bomo vse negativne dejavnike, ki se pojavljajo v skupinah ljudi brez portalov in za njih pripravili rešitev. Portal bomo izpopolnili tudi iz varnostnega vidika, ter predstavili načine uporabe, kako portal maksimalno izkoristimo.

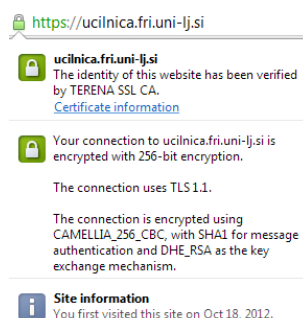
9.1 Varnost

Ker portali običajno vsebujejo tajne podatke, jih moramo skrbno varovati. Vsako zahtevo HTTP lahko s programom WireShark prestrežemo in pogledamo vsebino (Slika 9.1). Na portalu smo želeli poiskati vsebine, ki vsebujejo besedo »janez«. Tudi odgovor smo dobili v čistopisu, kar je za nas, kot lastnike podatkov, zelo nevarno. Podatke napadalec lahko vidi, lahko pa jih tudi spremeni. Nevarnost, ki preti, je tudi, da napadalec vidi, kakšne url naslove imamo. Če smo nepazljivi, lahko določen del portala ostane nezavarovan in ga napadalec obišče z neposrednim dostopom preko urlja v glavi zahteve.

Večja težava bi se pojavila, če SharePoint pri vpisu ne bi uporabljal podatkov iz imenika Active Directory in protokola NTLM (NT LAN Manager). Protokol NTLM poskrbi za autentikacijo, avtorizacijo in integriteto uporab-

```
{
  "queryxml": "<?xml version='1.0' encoding='utf-8' ?><QueryPacket xmlns='urn:Microsoft.Search.Query' Revision='1'><Query domain='QDomain'><Context><QueryText language='en-US' type='STRING'>janez</QueryText></Context><Range><StartAt>1</StartAt><Count>8</Count></Range><PreQuerySuggestions>true</PreQuerySuggestions><HighlightQuerySuggestions>false</HighlightQuerySuggestions></Query></QueryPacket>"
}
```

Slika 9.1: Prikaz nekriptirane vsebine paketa HTTP.



Slika 9.2: Primer podpisanega certifikata na spletni učilnici.

nikovih podatkov pri vpisu. Ostali podatki niso kriptirani avtomatsko.

Ko na portalu ustvarimo novo zbirko strani, se avtomatsko zaženejo procesi, ki dodajo spletno stran tudi v IIS. IIS je nadzorna plošča za urejanje Microsoftovega spletnega strežnika. V nadzorni plošči smo ustvarili novo lastno podpisan certifikat in ga namestili na našo spletno stran. Težava, ki se pojavi, je, da ko obiščemo spletno stran, nas opozori, da certifikat ni veljaven. Certifikat ni veljaven zaradi tega, ker ga nismo overili pri CA-ju (certificate authority). To načeloma pomeni, da lahko naše stranke dostopajo na napadalčevo spletno stran, ki je po obliki in stilu identična naši. Ne zavedajo pa se, da podatke pošiljajo napadalcu. Napadalec bi se moral v tem primeru potruditi, da bi manj večje uporabnike zvalil na svojo spletno stran. S podpisanim certifikatom zagotovimo strankam in uporabnikom portala, da je stran varna in da je v naši lasti. Primer strani s podpisanim certifikatom je Učilnica FRI (Slika 9.2). Zagotavlja nam, da bomo komunicirali preko varnega kanala TSL in da bomo komunicirali s pravim naslovnikom, Spletno učilnico FRI.

Pri overjanju certifikata moramo biti pozorni, da nam ga overi organizacija, ki ima zaupanje pri korenski CA organizaciji.

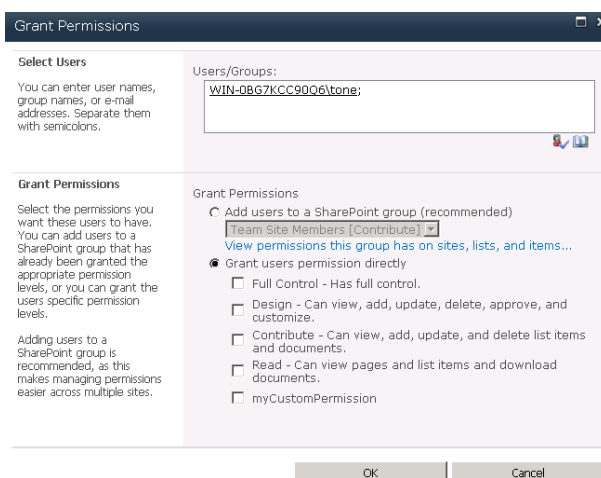
9.2 Pravice

Podjetja imajo načeloma hierarhično strukturo zaposlenih. To se navezuje tudi na pravice, kdo lahko vidi katere podatke. Portal nam ponuja kompleksni sistem urejanja pravic uporabnikov. Na portalu smo ustvarili pravilo, da določena skupina lahko samo pregleduje podatke. Pravilo ustvarimo na »Site Settings → Site permissions → Add new permissions level«. Ko smo ustvarili pravilo, ga dodelimo zeleni skupini. Uporabnike je priporočljivo združevati v skupine, da se izognemo dodeljevanju in nadaljnjemu administriranju pravil za vsakega posameznika. Trenutno pravilo velja za celotno zbirko strani in podstrani SharePoint aplikacije. V primeru, da želimo nekatere dele spletne strani podrobneje varnostno zavarovati, kliknemo na posamezen element, na primer dokumente v skupni rabi. Izberemo zavihek knjižnica (library) in nato kliknemo pravice knjižnice. V oknu izbrane knjižnice izberemo opcijo prenehanja dedovanja pravic in dodamo novo skupino z našimi pravicami (Slika 9.3). Slika 9.3 prikazuje tudi, kako določimo pravice skupine samo enemu uporabniku za posamezen element SharePoint portala.

Takšne pravice kot ima oseba za posamezen element na portalu upošteva tudi iskalnik. Če uporabnik nima dostopa do določenih vsebin, jih tudi v iskalnih rezultatih ne bo dobil. Da določen uporabnik lahko uporablja iskalnik, pa mu moramo že na nivoju strežnika AD dodeliti pravice uporabe storitve iskanja. Skupina se imenuje WSS_WPG.

9.3 Testni podatki

Da bi iskalnik lahko testirali, smo si morali domisliti način, kako bi poenostavljeno uvozili večje količine podatkov. Microsoft Visual Studio 2010 ima možnost uvoza dodatka knjižnice za obdelavo SharePointa 2010. Knjižnica



Slika 9.3: Okno za dodajanje pravic posamezni osebi ali skupini.

nam nudi veliko število objektov za komuniciranje in urejanje portala. Odločili smo se, da bomo podatke zbrali na takih mestih, kot bi jih potrebovali v realnem produkcijskem okolju. Nekateri podatki so shranjeni v datoteki na strežniku, nekaj jih je shranjenih v mapi Skupni dokumenti na portalu, približno 15 tisoč zapisov pa smo shranili tudi v zbirko, imenovano list. V list zbirkah običajno hranimo krajše in strukturirane zapise.

Za podatke smo izbrali kar vsebine internetnih strani. Program deluje tako, da mu podamo začetno stran. To stran razdeli na podstrani in strani, ki se povezujejo na druge strani. Na celotni spletni strani smo poiskali vse jač HTML elemente in iz njih pridobili vrednost atributa href. Da glavnih strani nismo preiskovali ponovno, smo vse domene shranili v podatkovno bazo. Podatke smo iz podatkovne baze vračali in shranjevali s shranjeno proceduro.

Pri prenosu podatkov na portal smo si pomagali z objekti iz knjižnice Microsoft.SharePoint. Najpogosteje uporabljena objekta pri našem programu sta bila SPSite in SPWeb (Primer 9.1). Celotna zgradba knjižnice temelji na hierarhiji (Slika 9.4) objektov. V našem primeru smo imeli samo en spletni strežnik, zato smo za dostop do spletne strani najprej uporabili razred



Slika 9.4: Hierarhija objektov knjižnice SharePoint [20].

SPSite. Spletni portal lahko vsebujejo več podstrani. Razdelimo jih lahko na oddelke v posamezni organizaciji. V programski kodi bi to realizirali tako, da bi iz objekta site vzeli lastnost AllWebs, ki predstavlja zbirko in iz te zbirke bi ponovno vzeli posamezno spletno podstran.

V programski kodi se nam ni bilo treba nikjer eksplicitno vpisovati v portal. Naša aplikacija podatke za vpis v portal pridobi avtomatsko glede na trenutno vpisanega uporabnika v operacijski sistem.

```

1 SPSite site = new SPSite(this.sharePointSite);
2 SPWeb web = site.RootWeb;
3 SPList list = web.Lists["myList"];
4 SPListItem listItem = list.AddItem();
5 listItem["Title"] = title;
6 listItem["URL"] = url; //. Substring(7, url.Length - 7);
7 listItem["Content"] = content;

```

```
8 listItem.Update();
```

Listing 9.1: Descriptive Caption Text

Posebej bi izpostavili razred SPQuery(Primer 9.2). Razred je zelo uporaben, ko želimo na portalu izdelati prirejen webpart, ki prikazuje samo določene zapise s portala. Obenem pa ne želimo vpisovati kakršnega koli besedila v iskalna polja. Procedura spodaj se poveže na spletno stran, poišče vse naloge trenutno prijavljenega uporabnika za današnji dan in jih izpiše.

```
1 public void ReadTasks()
2 {
3     using (SPSite site = new SPSite("https://localhost"))
4     {
5         using (SPWeb web = site.RootWeb)
6         {
7             SPList list = web.Lists["Tasks"];
8             String queryStr = "<OrderBy>" +
9                 "<FieldRef Name=\"Title\" Ascending=\"TRUE\"/>" +
10                "</OrderBy>" +
11                "<Where><And><Eq>" +
12                    "<FieldRef Name=\"AssignedTo\"/>" +
13                    "<Value Type=\"Integer\"> <UserID/> </Value>" +
14                    "</Eq><Eq>" +
15                    "<FieldRef Name=\"Created\"/>" +
16                    "<Value Type=\"DateTime\"><Today/></Value>" +
17                    "</Eq></And>" +
18                "</Where>";
19            SPQuery oQuery = new SPQuery();
20            oQuery.Query = queryStr;
21            SPListItemCollection collListItems=list.GetItems(oQuery);
22
23            foreach (SPListItem oListItem in collListItems)
24            {
25                Response.Write(SPEncode.HtmlEncode(oListItem["Title"].
26                    ToString() + " "
27                    + oListItem["Assigned To"].ToString()) + "<BR>");
28            }
29        }
30    }
```

```
29 }  
30 }
```

Listing 9.2: Descriptive Caption Text

9.4 Organizacija SharePointa

Pri načrtovanju portala je pomemben dejavnik tudi, kako bomo celoten portal organizirali [17]. Obstajati dve skrajnosti, kako lahko zadostimo našim potrebam. Prvi način je, da naredimo spletno aplikacijo in v njej naredimo podstrani za posamezen oddelek. Drugi način pa je, da naredimo spletne strani za vsak oddelek posebej. Prvi način je dobro uporabiti v primeru, da so skupine oseb in pravice na vseh podstraneh skoraj nespremenjene. Tako se izognemo precejšnjemu delu administriranja portala. Težave lahko nastanejo v primeru, če bi zaposleni iz prodaje želeli nastaviti pravice določenim strankam ali kreirati nov element na spletni strani. Tako bi moral eden od zaposlenih v prodaji postati administrator določenega dela portala. Če planiramo, da bomo imeli več ločenih enot, je boljša varianta, da naredimo za vsako enoto posamezno spletno stran in določimo odgovorno osebo. Primer, kdaj bi se odločili za drugo varianto, je na primer vodenje projektov. Pri vsakem projektu imamo odgovorno osebo. Podatkov glede projektov med seboj načeloma ne želimo mešati. Iskalnik imamo nastavljen tako, da preiskuje podatke samo na našem projektu. Ko pa projekt zaključimo, lahko preprosto zberemo vse podatke. Pri uporabi podstrani se nam lahko pojavijo težave s neveljavnimi povezavami na izbrisano vsebino.

9.5 BCS

Portal SharePoint nam ne ponuja samo kolaboracije med uporabniki. Omogoča nam tudi vpeljavo informacijskega sistema v portal. Tako lahko zaposleni brez preklapljanja aplikacij opravljajo svoje delo. Uporabnikom omogoča, da iščejo po zunanjih podatkovnih bazah [18]. S pojmom zunanje podat-

kovne baze označujemo podatke, ki niso del SharePointa. Storitve lahko izkoristimo tudi takrat, ko se naša stranka odloči, da bo svoje poslovanje preselila na portal. Tako nam ni treba nikamor kopirati podatkov, ampak samo nastavimo pot, kjer so. Za demonstracijo delovanja smo vzeli podatkovno bazo NorthWind in ji dodali še nekaj tabel z našimi podatki. Na portalu smo morali najprej ustvariti storitev BCS. Določiti smo ji morali bazo, kam bo shranjevala nastavitve, meta podatke o podatkih in Application pool. Application pool je metoda ločevanja procesov. Vsaki spletni strani ali storitvi lahko v nadzorni plošči Microsoftovega strežnika naredimo svoj Application pool in tako zmanjšamo možnost prenehanja delovanja celotnega strežnika. Če se dogodi napaka v enem izmed aplikacijskih področij, se ta ustavi in ponovno zažene, ne da bi to motilo ostale aplikacije, ki se izvajajo vzporedno. Sledi izdelava modela, preko katerega storitev in uporabniki dostopajo do podatkov. Model s programom Designer dodamo na portal. V tej stopnji BCS že nemoteno deluje, ampak iskalni rezultati ne prikazujejo berljivih podatkov za uporabnika. Vsakemu izdelanemu modelu moramo narediti tudi prikazno stran.

9.6 Nastavitve pregledovalnika

Ena izmed pomembnejših komponent je pregledovalnik. Če želimo, da so nam vsi podatki dostopni, moramo največ pozornosti posvetiti ravno pregledovalniku. Microsoft nam ponuja možnost pregledovanja po veliko različnih virih in vrstah datotek. Nastavili in pregledali smo vse razpoložljive vire. Z BCS smo se povezali na zunanjo podatkovno bazo, preiskali smo strani in zapise na portalu, preučili smo možnost pregledovanja strani z drugih domen in nastavili pregledovalnik na datoteko za izmenjavo na strežniku. Pregledovanje zunanjih spletnih strani nam koristi v primeru, če nas zanimajo podatki druge organizacije, a nimamo njihove podatkovne baze. Mi smo nastavili pregledovanje na spletni strani podjetja Grifon in Castel. Pregledovalnik je pregledal vse strani na domeni castel.si in grifon.si. Dovolili smo mu

Search Query Text	Number of Queries	Percentage of Overall
1 task1	5	12.20 %
2 slana	3	7.32 %
3 janez	2	4.88 %
4 samsung	2	4.88 %
5 select*	2	4.88 %
6 StreamInsight	2	4.88 %

Slika 9.5: Prikaz števila in iskalnih nizov.

tudi en preskok na druge spletne strani. Če spletna stran vsebuje povezavo na simobil.si, je pregledovalnik pregledal tudi to. Pregledovanje datotek na strežniku je zelo pomemben element. Tak način pregledovanja je koristen iz več razlogov. Datotek nam ni treba prenašati na portal in tudi varnostne kopije portala so bolj obvladljive. Obstaja tudi slabost, da je treba skrbeti za dostop do podatkov na dveh koncih, na portalu in na strežniku. Pregledovalniku lahko določimo tudi pravila. Pravila omejujejo, katere podatke lahko iskalnik pregleduje. Nastavili smo, da iskalnik ne sme pregledovati vsebin, ki jih pridobi z zahtevo »http:/*print=true«. SharePoint nam dovoljuje vnesti do 100 pravil, a se moramo zavedati, da pravila povečujejo čas pregledovanja. Posebnost, ki jo omogoča portal, je tudi pregledovanje po datotekah s končnico .pdf. Datoteke pdf so sestavljene iz strani, ki predstavljajo slike. Da je pregledovalnik lahko razčlenil vsebino, smo morali namestiti Adobe PDF iFilter. Dodatek uporablja optično prepoznavanje znakov (ORC - Optical character recognition).

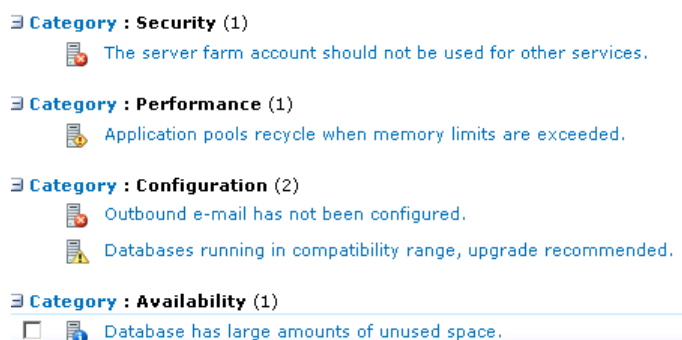
9.7 Uporaba iskalnika

Uporaba iskalnika je zelo intuitivna zadeva. V iskalno vrstico napišemo iskalni niz in počakamo na rezultat. Velika verjetnost je, da bomo tudi dobili prave iskane rezultate. Na točnost rezultatov vpliva tudi številka, koliko iskalnih poizvedb smo imeli in kakšni so bili iskalni nizi (Slika 9.5).

Če je ima portal manjše število iskalnih poizvedb in veliko število dokumentov, se verjetnost, da bomo našli, kar iščemo, zmanjša. V takih primerih je dobro poznati trike posameznega iskalnika (Google, SharePoint, Bing ...).

SharePointov iskalnik podpira logične operatorje, nadomestne znake (wild-cards), poizvedbe na specifične lastnosti in navadne tekstovne poizvedbe. Primer poizvedbe, ki vsebuje specifične lastnosti, je »ALL(Hola aus Barcelona) Author:administrator«. Poizvedba vrne rezultate, ki vsebujejo vse besede, njihov avtor pa je administrator. Specifične lastnosti zapisov lahko izbiramo, če ob vnosnem polju za poizvedbo kliknemo povezavo napredno. V logične operatorje spadajo AND, OR in NOT. AND in OR že sama opisujeta, kakšna je njihova vloga in uporaba. NOT operator uporabljamo tako, da ga postavimo pred besedo in iskalnik bo izločil vse zadetke, ki vsebujejo tak niz. Če želimo izločiti več besed, uporabimo ukaz NONE(niz znak). Pogosto uporabljeni ukazi so tudi NEAR in WORDS. Ukaz NEAR deluje podobno kot operator AND. Primer ukaza: NEAR(pogodba odobrena). Operator AND bi nam vrnil vse zadetke, ki bi vsebovali besedo pogodba in odobrena, dobili bi tudi takšne, ki imajo besedo pogodba na začetku dokumenta, odobrena pa bi bila na koncu. Ukaz NEAR pa nam vrne samo tiste rezultate, ki imajo manj kot osem terminov med besedama pogodba in odobrena. Iskalne rezultate si močno izboljšamo tudi z uporabo ukaza WORDS. Ukaz WORDS deluje podobno kot logični operator OR. Razlika je v rangiranju rezultatov. Primer uporabe ukaza: WORDS(marljiv delaven). Če imamo dokument, ki vsebuje pet besed marljiv in pet besed delaven, bi ukaz WORDS vrnil deset ponovitev. Pri uporabi logičnega operatorja OR SharePoint interpretira rezultat, da je našel pet ponovitev marljiv in pet ponovitev delaven. Iskalni zadetki so razvrščeni padajoče po številu ponovitev.

Preden potrdimo iskalni niz, lahko nastavimo tudi področja, po katerih želimo iskati. Področja (scopes) naredimo v centralni konzoli. Vsakemu področju moramo nastaviti, kateri nabor vsebin bo zajemal. Iskalnik omogoča tudi podajanje predlogov. Predloge se mora iskalnik naučiti glede na naše predhodne iskalne vnose. Ko nam strežnik vrne rezultate, jih lahko še izboljšamo. Iskalnik nam ponudi združene lastnosti iskanih zadetkov.



Slika 9.6: Prikaz opozoril in napak v sistemu.

9.8 Ukazna vrstica

Pri upravljanju celotnega portala imamo na voljo več možnosti. Uporabljamo lahko administrativno centralo ali pa ukazno vrstico PowerShell. Veliko večino našega dela smo opravili preko uporabniškega vmesnika. Uporabniški vmesnik je zelo izpopolnjen. Dostopali in nastavili smo vse želene nastavitve. Velika potreba po ukazni vrstici se bi nam pojavila, ko bi želeli narediti ponavljajoče se aktivnosti ali pa nastaviti zelo podrobne nastavitve portala. Primer nastavitve je poimenovanje administrativne podatkovne baze ob instalaciji portala.

9.9 Posodobitev podatkovne baze

Ker smo v času razvoja na portal namestili kar nekaj novih kar nekaj novih funkcionalnosti, nas je sistem (Slika 9.6) opozoril, da je treba posodobiti podatkovne baze. Z ukazom `stsadm.exe` (Primer 9.3) smo se najprej prepričali, katera podatkovna baza je zastarela. Odgovor smo dobili v XML obliki. Vseboval je navodila, na kateri podatkovni bazi moramo izvesti določeno akcijo. V našem primeru je bila to konfiguracijska podatkovna baza. Nedelovanje podatkovne baze `SPConfigurationDatabase` se je odražalo tudi tako, da na portalu nismo mogli shraniti nobene nastavitve.

```
1 stsadm.exe -o localupgradestatus
2
3 <objects>
4   <object>
5     <name>SharePoint_Config_b9eebe23-7ed6-4a44-a905-456cd644a058</name>
6     <type>Microsoft.SharePoint.Administration.SPConfigurationDatabase</type>
7     <level>0</level>
8     <status>Needs Upgrade</status>
9   </object>
10 </objects>
11
12 psconfig -cmd upgrade -inplace b2b
```

Listing 9.3: Pododobitev podatkovne baze.

Poglavje 10

Zaključek

V praktičnem delu diplomske naloge smo izdelali manjši sistem, s katerim bi lahko podprli delovanje organizacije z do petdeset zaposlenih. Platforma SharePoint Foundation nam skupaj z iskalnim strežnikom Search Server Express ponuja veliko prednost na področju izmenjave dokumentov in povezovanja dela zaposlenih. Pri testiranju izdelka smo nekaj časa namenili tudi izdelavi informacijskega sistema na portalu SharePoint. Tukaj so se pojavile največje razlike. Brezplačna verzija Foundation programerjem ne omogoča takšnih možnosti izdelovanja dodatkov na portalu kot ostale plačljive verzije. Najbolj je bila pogrešana storitev InfoPath, ki nam v verziji SharePoint Server omogoča izdelavo kompleksnih form za vnos podatkov v sistem. Pogrešali smo tudi gradnike za prikazovanje podatkov s tabel v podatkovni bazi. Ti so nam namreč na voljo samo v verziji Server. Slabost portala je tudi prenos na novejšo verzijo in posodobitve. Iz neuradnih virov smo izvedeli, da portala ni priporočljivo nadgrajevati, ker lahko odpove. Tudi migracije na novo verzijo niso preproste. Microsoft je v času nastajanja diplomske naloge izdal novo verzijo portala, imenovano SharePoint 2013.

S portalom in iskalnikom smo na splošno zelo zadovoljni. Portal bi lahko primerjali z nekaterimi odprtokodnimi CMS-ji, kot sta WordPress in Joomla. Težje bi našli konkurenco v iskalnikih.

Literatura

- [1] Chris Geier, Cathy Dew, Becky Bertram, Raymond Mitchell, Wes Preston, Ken Schaefer, Andrew Clark, SharePoint 2010 Six-in-One, Indianapolis: Wiley Publishing, 2011, pogl. 1.
- [2] Josh Noble, Robert Piddocke, Dan Bakmand-Mikalski, Pro SharePoint 2010 Search, New York: Apress, 2011.
- [3] (2012) Architecture design for SharePoint 2013 IT pros (Topologies, Search architectures, Enterprise search architectures). Dostopno na: <http://technet.microsoft.com/en-us/sharepoint/fp123594.aspx>.
- [4] (2012) Microsoft Visual Studio. Dostopno na: http://en.wikipedia.org/wiki/Microsoft_Visual_Studio.
- [5] (2012) Licenciranje in funkcionalnosti verzij. Dostopno na: <http://sharepoint.microsoft.com/en-us/buy/Pages/default.aspx>.
- [6] (2012) Različice iskalnikov. Dostopno na: <http://zoom.it/So9H#full>.
- [7] (2012) Lastnosti iskalnikov SharePoint. Dostopno na: <http://blogs.technet.com/b/vedant/archive/2009/10/23/search-technologies-for-sharepoint-2010-products.aspx>.
- [8] (2012) Arhitektura iskalnikov SharePoint. Dostopno na: <http://msdn.microsoft.com/en-us/library/ms570748%28v=office.12%29.aspx>.
- [9] (2012) Vrste omrežji. Dostopno na: <http://www.hcidata.info/inet.htm>.

-
- [10] (2012) Microsoft.NET framework. Dostopno na: http://en.wikipedia.org/wiki/.NET_Framework.
- [11] (2012) Podatkovne baze portala SharePoint. Dostopno na: <http://blogs.technet.com/b/wbaer/archive/2009/11/30/introduction-to-the-microsoft-sharepoint-sharepoint-2010-database-layer.aspx>.
- [12] (2012) Spletne storitve portala SharePoint. Dostopno na: <http://msdn.microsoft.com/en-us/library/ee538665.aspx>.
- [13] (2012) Priporočila pri instalaciji SharePoint portala. Dostopno na: <http://goo.gl/WxLNf>. Poglavje: Physical architecture.
- [14] (2012) Izračun velikosti podatkovnega prostora. Dostopno na: <http://technet.microsoft.com/en-us/library/cc298801.aspx>.
- [15] (2012) Testi in postavitve arhitekture sistema. Dostopno na: <http://technet.microsoft.com/en-us/library/gg750251.aspx>.
- [16] (2012) Strukture sistemov. Dostopno na: <http://www.slideshare.net/itsec/sharepoint-topology>.
- [17] (2012) Plan for site collections. Dostopno na: [http://technet.microsoft.com/en-us/library/ff427197\(v=office.12\).aspx](http://technet.microsoft.com/en-us/library/ff427197(v=office.12).aspx).
- [18] (2012) Overview of Business Connectivity Services. Dostopno na: [http://technet.microsoft.com/en-us/library/ee661740\(v=office.15\).aspx](http://technet.microsoft.com/en-us/library/ee661740(v=office.15).aspx).
- [19] (2012) Sharepoint platform services. Dostopno na: <http://goo.gl/eSwxA>.
- [20] (2012) Slika hierarhije objektov v knjižnici SharePoint, Slika 9. Dostopno na: <http://eprints.fri.uni-lj.si/1590/1/Rupnik1.pdf>.