



UPPSALA
UNIVERSITET

UPTEC IT 14 006

Examensarbete 30 hp
April 2014

Vidareutveckling av sökmotor i EPiServer CMS

Isabelle Håff Hedung



UPPSALA
UNIVERSITET

**Teknisk- naturvetenskaplig fakultet
UTH-enheten**

Besöksadress:
Ångströmlaboratoriet
Lägerhyddsvägen 1
Hus 4, Plan 0

Postadress:
Box 536
751 21 Uppsala

Telefon:
018 – 471 30 03

Telefax:
018 – 471 30 00

Hemsida:
<http://www.teknat.uu.se/student>

Abstract

Improvement of the search functionality in EPiServer CMS

Isabelle Håff Hedung

EPiServer is a content management system where users can publish information on a website by using a web browser. EPiServer has a built-in search engine that is very simple. Many customers using EPiServer web publishing system is in need of more functionality than is offered by EPiServer's built-in search engine. The purpose of this thesis is to develop additional functionality extends that the built-in search engine in EPiServer. The assignment was also to identify what need to be developed in order to give customers a better opportunity to help their visitors find the right information on their website.

Since this thesis was started in 2005, a number of commercial search engine products entered the market and it has become more common for companies to invest in these products instead of using the simple search engine that's built into EPiServer. These products have many of the features that have been developed in this thesis. Another crucial difference between EPiServer's built in engine and other products is the method used to index content. The approach that the commercial search engines use will make the relevance of search results to be much better.

The product that was developed in this thesis fulfilled the demands of the customers well, and the result of this thesis has been distributed to several of the company's customers. However, in recent years it has been proven to be difficult to compete with the commercial search engines.

Handledare: Erik Nettelbrandt
Ämnesgranskare: Arnold Pears
Examinator: Lars-Åke Nordén
ISSN: 1401-5749, UPTEC IT 14 006
Sponsor: Avantime Group

Tryckt av: Reprocentralen ITC

SAMMANFATTNING

EPiServer är ett webbpubliceringssystem där användaren kan publicera information på en webbplats direkt via en webbläsare. I EPiServer finns en inbyggd sökmotor som är väldigt enkel. Många kunder som använder EPiServer som webbpubliceringssystem är i behov av ytterligare funktionalitet än vad som erbjuds av EPiServers inbyggda sökmotor. Syftet med detta arbete är att utveckla funktionalitet som bygger ut den inbyggda sökmotorn i EPiServer. Uppgiften har även bestått av att identifiera de delar som behöver utvecklas för att ge kunderna bättre möjlighet att kunna hjälpa sina besökare att hitta rätt information.

Sedan detta arbete inleddes under 2005 har ett antal kommersiella sökmotorprodukter kommit in på marknaden och det har blivit allt vanligare att företag investerar i dessa produkter istället för att använda den enklare sökmotor som finns inbyggd i EPiServer. Dessa produkter har många av de funktioner som tagits fram i detta arbete.

En annan avgörande skillnad mellan EPiServers inbyggda sökmotor och de övriga produkterna är metoden som används för att indexera innehåller. Den metod som de kommersiella sökmotorerna använder gör att relevansen i sökresultatet blir mycket bättre.

Produkten som togs fram i detta arbete uppfyllde önskemålen från kunderna väl och resultatet av arbetet har distribuerats till flertalet av företagets kunder. Dock har det under senare åren har det visat sig vara svåra att konkurrera med de kommersiella produkterna.

FÖRORD

Denna rapport är resultatet av examensarbete för Civilingenjörsutbildningen inom IT programmet på Uppsala universitet. Arbetet påbörjades under våren 2005.

Uppdragsgivare för examensarbetet är Avantime AB. Avantime vill ha en komplett sökmotorlösning att erbjuda sina kunder. Den sökmotor som finns tillgänglig idag har brister och behöver kompletteras med ytterligare funktionalitet.

När detta arbete gjordes under 2005 låg det i framkant och resultatet av arbetet har även distribuerats till flera av företagets kunder. Några av dem använder produkten än idag.

Sedan 2005 har ett antal kommersiella sökmotorprodukter kommit in på marknaden som det visat sig svårt att konkurrera med. Bland de främsta är Google Search Appliance (GSA), SiteSeeker samt Microsofts sökmotor som finns inbyggd i Sharepoint-plattformen.

INNEHÅLL

Sammanfattning	1
Förord	2
Inledning	5
Om EPiServer	5
Om avantime	5
Bakgrund.....	5
Syfte	5
Problemformulering	5
Analys	7
EPiServer	7
EPiServers inbyggda sökmotor	7
Relevansen i sökresultatet.....	8
Sökordslogg	9
Klicklogg.....	10
Rättstavning och synonymer	11
Avancerad sökning.....	11
Markering av sökt ord i sökresultaten	12
Avgränsad sökning	12
Söka i datumintervall	12
Uppdelning av sökresultatet.....	12
Sortering av sökresultatet	12
Olika typer av dokument	12
Metod	13
Systembeskrivning.....	14
Redigeraläge i EPiServer	14

Administrationsläge i EPiServer	16
Funktionalitet på en söksida	17
Resultat.....	18
Resultat	18
Diskussion	18
Slutsats.....	18
Referenser	19

INLEDNING

OM EPISERVER

EPiServer är ett webbpubliceringssystem där användaren kan publicera information på en webbplats utan kunskap om tillverkning av webbsidor. Arbetet sker från en webbläsare och kan jämföras med enkel ordbehandling.

OM AVANTIME

Avantime AB utvecklar IT-lösningar till företag. En stor del av verksamheten kretsar kring utveckling av webbplatser och webbapplikationer baserade på EPiServer. Företaget har även utvecklat flertalet tilläggsprodukter till EPiServer.

BAKGRUND

I EPiServer finns en inbyggd sökmotor som är väldigt enkelt. Den indexerar alla ord i alla webbsidor i en tabell i databasen tillsammans med en räknare på antalet förekomster av ordet på webbsidan. När en besökare gör en sökning returneras de sidor där ordet finns indexerat. EPiServer gör en relevanssortering som är baserad på antalet förekomster av ordet på webbsidan.

SYFTE

Syftet med detta arbete är att utveckla funktionalitet som bygger ut EPiServers inbyggda sökmotorn för att ge besökare ett bättre verktyg för att hitta det som de söker efter och att ge redaktörer möjlighet att kunna påverka och förbättra sökresultatlistan.

Målet är att bygga ett så enkelt och användarvänligt gränssnitt för redaktören trots att det är avancerad och komplex funktionalitet som ligger bakom. Besökaren skall inte heller märka något av den komplexiteten bakom sökmotorn.

Uppgiften består i att identifiera de delar som behöver utvecklas för att uppnå ovanstående mål.

PROBLEMFÖRMULERING

Tekniken för att publicera information på nätet har blivit bättre och bättre under det senaste årtiondet vilket resulterar i att företag och organisationers webbplatser ökar i storlek. Alltmer information läggs ut på företagets/organisationens hemsidor och många företag har i dag sin webbplats som främsta informationskanal. Ofta hänvisas till företagets hemsida i reklamfilmer, broschyrer och även på telefon. Enligt min mening är det en bra utveckling då informationen är tillgänglig för alla, dygnet runt.

Med en ökande mängd information på en webbplats kan det också följa problem. Det är svårare att hitta rätt till den information som söks efter i den stora mängden av information. En ovan besökare kan ha svårare att hitta rätt. Många webbplatser är indelade i några få menyer, under dessa menyer kan otaligt mycket information finnas. Hur ska en besökare veta vilken menyrybrik just den informationen han/hon

eftersöker finns? Krav på att det ska finnas en sökmotor kopplad till en hemsida är idag självklart. Om sökmotorn inte ger det resultat som förväntas kan det vara frustrerande för besökaren.

ANALYS

EPISERVER

EPIserver är ett s.k. Content Management System (CMS) eller webbpublicerings verktyg. Ett CMS är ett datorsystem som gör det möjligt för användare att publicera information på en hemsida utan någon teknisk kunskap om tillverkning av webbsidor. Arbetet sker från en webbläsare och kan jämföras med enkel ordbehandling.

Det finns tre olika ingångar till en EPIserver site, först och främst den vanliga hemsidan som är öppen för alla besökare. En vanlig besökare märker inget av att det är ett CMS som ligger bakom hemsidan.

Den andra ingången är Edit-läget som webbredaktörer har tillgång till. Här kan nya sidor skapas, befintliga uppdateras med ny information och menyn kan struktureras om genom enkel drag-and-drop. Detta är några av de grundläggande funktionerna i redigera läget i EPIserver. Utöver det finns det många fler avancerade funktioner såsom behörighetshantering, dokumenthantering, arbetsflöden för publicering av sidor osv.

Den tredje ingången är Admin-läget som en administratör har tillgång till. Här görs mer avancerade inställningar och konfigurering i EPIserver.

EPIserver är baserat på Microsoft ASP.Net Plattform, vanligtvis med Microsoft SQL Server som databas i botten. All utveckling sker i C#, Transact-SQL, samt HTML, CSS och javascript i utvecklingsmiljön Microsoft Visual Studio.

EPISERVERS INBYGGDA SÖKMOTOR

I EPIserver finns en inbyggd sökmotor som indexerar alla ord som ligger på en hemsida. Varje ord läggs i en tabell tillsammans med en räknare på antalet förekomster på en viss sida. När en sökning görs returneras de sidor där ordet förekommer. Resultatet sorteras baserat på antalet förekomster på sidan.

I EPIservers inbyggda sökmotor har ett antal punkter som kräver förbättring har identifierats. Dessa punkter har sammanställts och analyserats var och en för sig. Här följer dessa punkter:

- Relevansen i sökresultatet
- Sökordslogg
- Klicklogg
- Avancerad sökning
- Markering av sökt ord i sökresultaten
- Rättstavning och synonymer
- Avgränsad sökning
- Söka i datumintervall
- Uppdelning av sökresultatet
- Sortering av sökresultatet
- Olika typer av dokument

Sök

[Avancerad sök »](#) [Visa söktips »](#)

Sökresultat

Din sökning på "bytestorget" resulterade i 4 träff(ar).

Sortera efter

Relevans Kategori Datum

♦ [Bytestorget nedlagt tillsvidare](#) - Övrigt 2010-11-05

Vi vill härmed meddela att Bytestorget är nedlagt tillsvidare på grund av att det har haft tekniska brister. Under våren nästa år kommer vi att utveckla vår hemsida och därmed hitta en ny bättre...

[Hem](#) / [Press och Media](#) / [Senaste nytt](#)

♦ [Hyrestorget](#) - Hyrestorget 2010-11-05

På Hyres & Bytestorget hittar du information om lediga lägenheter och bostäder i Sigtuna, Märsta, Valsta och Rosersberg

[Hem](#)

♦ [Bytestorget på hemsidan](#) - Övrigt 2010-10-28

Sigtunahems bytestorg ligger för tillfället nere. Vi arbetar med att lokalisera felet.

[Hem](#) / [Press och Media](#) / [Senaste nytt](#)

♦ [Information](#) - Hyrestorget 2011-02-02

För att kunna söka bostad och lägenhet på Hyrestorget måste man registrera sig, läs mer om processen här.

[Hem](#) / [Hyrestorget](#)

Figur: Exempel på ett sökresultat

RELEVANSEN I SÖKRESULTATET

I den inbyggda sökmotorn finns ingen möjlighet för en redaktör att påverka relevansen i sökresultatet. Relevansen styrs enbart på antalet förekomster av sökordet på sidan. I vissa fall kan det finnas önskemål om att en specifik sida ska presenteras högre upp i sökresultatet även om ordet som eftersöks kanske förekommer fler gånger på andra sidor. Orsaker till det kan vara att det finnas mer aktuella sidor som presenterar informationen på ett bättre sätt.

För att ge redaktören möjlighet att styra relevansen har två olika lösningar implementerats. Den första lösningen ger redaktören möjlighet att kunna påverka relevansen genom att "lura" sökmotorn. På varje sida på hemsidan kan redaktören ange ett par extra nyckelord som ska kopplas till sidan. Dessa sökord döljs från en vanlig besökare, men visas på ett sådant sätt att sökmotorn tror att det är som vilket ord som helst på sidan. Genom att repetera nyckelorden flera gånger för sökmotorn så kommer relevansen på dessa sidor påverkas när en besökare söker på angivna nyckelord.

Den andra lösningen är att så kallade utvalda träffar presenteras ovanför resten av sökresultatet. Det innebär att för varje sökning som besökaren gör så görs egentligen två sökningar mot sökmotorn. Den ena sökningen är en vanlig sökning mot hela hemsidan medan den andra sökningen enbart görs bland de utvalda träffarna. Om träff fås på en utvald träff så presenteras den ovanför resten av sökresultatet på ett sådant sätt att den ska fånga besökarens blick. På så vis har redaktören stor möjlighet att hjälpa besökaren till rätt sida.

SÖKORDSLOGG

När en besökare gör en sökning görs med EPiServers vanliga sökmotor görs ingen loggning. Det är önskvärt för en redaktör att kunna se vad besökare på hemsidan söker efter för information. På så vis kan hemsidan förbättras och anpassas för att tillgängliggöra information som besökarna eftersöker. Det kan även vara intressant att se vad besökare söker efter inom ett visst tidsintervall.

Så fort någon gör en sökning på hemsidan så sparas datum och tidpunkt för sökningen tillsammans med sökorden i en egen tabell i databasen. En redaktör har sedan möjlighet att via ett användarvänligt gränssnitt gå in och titta på dessa listor.



Rank	Sökt fras	Sökningar	Träffar	
1	bytestorget	3	4	+
2	tillval skåpluckor	2	0	+
3	tingvalla	1	14	+
4	trollbergsvägen	1	2	+
5	tvbox	1	0	+
6	utveckling norrbacka	1	0	+
7	Valsta kontor	1	1	+
8	vikingavägen	1	4	+
9	vårberg	1	0	+
10	västra ban gatan	1	0	+

Figur: Sökordslogg

Listorna presenteras under rubriken "Sökta ord" i ett administrationsgränssnitt i EPiServer. Här kan redaktören ställa in ett antal kriterier för att plockar fram data ut sökloggen. De inställningar som kan göras är följande: möjlighet att välja hur många sökningar som skall visas, möjlighet att se hela fraser eller

separata ord, möjlighet att välja att se totalt antal sökningar oberoende av tid, att se sökningar som gjorts under den senaste veckan, den senaste månaden eller att helt fritt välja startdatum och slutdatum. Som standard hämtas totalt antal sökta fraser oberoende av tid. Redaktören kan sedan ändra inställningar.

Resultatet presenteras i en lista med fyra kolumner. Den första kolumnen är rank som bestäms av hur många sökningar som gjorts på resp. ord/fras. Nästa kolumn är ordet/frasen, därefter kommer antal sökningar som gjorts på ett och samma ord/fras och sist presenteras om ordet/frasen finns i synonym- eller rättstavningsordlistan. Alla kolumner utom den sista är sorteringsbara. Genom att klicka på kolumnens rubrik sorteras listan efter innehållet i den kolumnen. Genom att klicka flera gånger görs en omvänd sortering.

Varje ord/fras i den andra kolumnen är länkad direkt till en sökning av ordet/frasen där sökresultatet presenteras tillsammans med hur många klick som gjorts på respektive sökresultat. (Se nästa avsnitt för att läsa mer om klickloggen.)

KLICKLOGG

I den inbyggda sökmotorn finns ingen funktionalitet för att se vad en besökare får för resultat samt vad besökaren väljer att läsa för information. Det är önskvärt att kunna spåra vilka sidor en besökare väljer att klicka på i ett sökresultat. På så vis kan en redaktör se om relevansen i sökresultatet är bra och om det är lätt för besökaren att hitta det som eftersöks.

I sökordsloggen (beskriven ovan) kan man från sökordet klicka sig vidare till en klicklogg för sökresultatet. Där kan man se hur många som valt att klicka på resp sökträff.

KLICKLOGG		«
Sökfras: jobb		
Rank	Träff	Antal klick
99%	2000 Läs mer om Sigtuna kommun och Sigtunahem – platsen man vill bo, leva och verka inom nu och i framtiden	8
49%	Allmännyttan1 Läs mer om allmännyttan och välfärdsbygget mellan åren 1850-1925	0
49%	Märstagårdarna Läs mer om Märsta, Märstagårdarna, Märstabostäder, Sigtuna kommun och Sigtunahem	8
49%	1960 Läs om Märsta på 1960-talet hos Sigtunahem - ledande inom bostäder och lägenheter i Sigtuna, Märsta, Valsta och Rosersberg	1
49%	Jobba hos oss AB SigtunaHem är ett bostadsföretag som erbjuder attraktiva hyreslägenheter i Sigtuna kommun. Vi äger och förvaltar cirka 5 300 lägenheter där det bor cirka 9 000 människor, det vill säga var fjärde Sigtunabo. Vi är ett 80-tal personer som arbetar här och vår vision är att ha Sveriges mest nöjda hyresgäster. Dit vill vi nå genom att arbeta enligt följande värderingar; Vi arbetar som ett lag! Vi...	0
49%	2000A Läs mer om framtiden för Sigtuna kommun och Sigtunahem	0
49%	Nya Valsta Centrum Sigtunahem och Sigtuna kommun äger gemensamt (49/51) Valsta Centrum från 1 november 2010. Nu vill vi skapa en handels- och mötesplats för i första hand Valsta- och Steningeborna att vara stolta över. Som man väljer för att kompletteringshandla på väg hem från arbetet. Som man går till för att ta en kopp kaffe och prata bort en stund med vänner och bekanta. Som man aktivt väljer var i kommunen...	0
49%	Nya Valsta Centrum Från första november tar vi och Sigtuna kommun över Valsta Centrum. Följ med utveckling av Nya Valsta Centrum här på vår hemsida och Facebook.	0
		« Tillbaka

Figur: Klicklogg

RÄTTSTAVNING OCH SYNONYMER

Den inbyggda sökmotorn har ingen rättstavningsfunktion. Önskvärt vore att sökmotorn automatiskt kände av om ett ord är felstavat och istället ger träffar på det rättstavade ordet.

Efter att ha undersökt möjligheterna kring att bygga någon typ av automatiskt rättstavningsfunktion har det visat sig att det skulle vara väldigt komplext och skulle kräva mycket mer arbete än vad som kan innefattas i detta examensarbete. Därför har en lösning gjorts som innebär att en redaktör kan mata in en ordlista till sökmotorn samt ange vanliga felstavningar på dessa ord. Sökmotorn ger sedan förslag på annan stavning på ord som besökaren stavat fel med en länk direkt till en sökning på det rättstavade ordet. Samma sak gäller för synonymer.

AVANCERAD SÖKNING

I EPiServer finns redan idag möjlighet att göra avancerad sökning med booleska operatörer. Besökaren kan välja att ange + respektive – framför varje ord i sökfrasen för att välja in eller bort ord som skall ingå i sökresultatet. Dock är det standard att om inga booleska operatörer anges så blir det en sk. ”eller-sökning” vilket innebär att det räcker om något av orden som eftersöks finns med i sökresultatet.

Detta har gjorts om så att det istället automatiskt blir ”och-sökning” vilket innebär att alla ord som eftersöks måste finnas med i sökresultatet.

MARKERING AV SÖKT ORD I SÖKRESULTATEN

I den inbyggda sökmotorn görs ingen markering av det sökta ordet/orden i sökresultatet. För att underlätta för besökaren att hitta rätt i sökresultatet har en lösning implementerats som gör att det sökta ordet/orden markeras i sökträffarna.

AVGRÄNSAD SÖKNING

I EPiServers inbyggda sökmotor finns ingen möjlighet att välja att söka i delar av en webbplats (kategorier). Det är önskvärt att ha en sådan funktion samt att en redaktör ska kunna bygga upp vilka kategorier som ska finnas på hemsidan.

SÖKA I DATUMINTERVALL

I EPiServer finns inget enkelt sätt att filtrera sökresultatet på datum. Det kan vara önskvärt för en besökare att välja att enbart söka i sidor som tex publicerats under det senaste året.

UPPDELNING AV SÖKRESULTATET

I EPiServer finns inte heller någon möjlighet för en redaktör att styra hur många sökträffar som skall visas per sida i sökresultatet. Det är önskvärt att en redaktör ska kunna styra det på ett enkelt sätt. Idag måste en utvecklare göra det programmatiskt.

SORTERING AV SÖKRESULTATET

I EPiServer finns redan möjlighet att kunna sortera sökresultatet beroende på relevans eller datum men funktionen behöver sammanställas och paketeras på ett bättre sätt. Idag krävs att en utvecklare gör detta programmatiskt.

OLIKA TYPER AV DOKUMENT

I EPiServer finns inbyggd funktionalitet för att enbart söka en viss typ av dokument (tex Word, Excel, PowerPoint, PDF etc.). Idag krävs att en utvecklare går in och konfigurerar detta programmatiskt, denna funktionalitet ska sammanställas, dokumenteras och paketeras på ett bättre sätt.

METOD

I början av detta projekt fanns ett antal önskemål från Avantime på hur de ville förädla den produkt de tidigare använde som sökmotor till sina kunder. Dessa önskemål sammanfattades i en kravspecifikation. Var och en av dessa önskemål utreddes därefter mer noggrant och lösningsförslag togs fram och i vissa fall implementerades direkt. I andra fall krävdes ytterligare iterationer med kravställaren på Avantime.

En förstudie kring utveckling i EPiServer, Microsoft Visual Studio samt Microsoft SQL Server var också nödvändig innan utvecklingsarbetet kunde påbörjas.

SYSTEMBESKRIVNING

REDIGERALÄGE I EPISERVER

Meny

- Sökta ord
- Klicklogg
- Sökning utan träffar
- Synonymordlista
- Rättstavningsordlista
- Inställningar

MENY

Menyn ligger som en flik i det vänstra fältet i redigeringsläget i EPiServer. Menyn består av länkar till de olika delarna av redigeringsgränssnittet. Länkarna öppnas i Åtgärdsfönstret.

The screenshot shows the EPiServer 4 editor interface. On the left, there is a navigation menu with options like 'Struktur', 'Favoriter', 'Aktiviteter', 'AvanSearch', and 'AvanTools'. The main content area displays a page for 'SigtunaHem' with a search bar and a 'SÖKORDSLOGG / SÖKTA ORD' section. The search results are displayed in a table with columns for Rank, Sökt fras, Sökningar, and Träffar.



Rank	Sökt fras	Sökningar	Träffar
1	bytestorget	3	4
2	tilval skåpluckor	2	0
3	tingvalla	1	14
4	trolbergsvägen	1	2
5	tvbox	1	0
6	utveckling norrbacka	1	0
7	Valsta kontor	1	1
8	vikingavägen	1	4
9	västra ban gatan	1	0
10	bytt lägenhet	1	1

SÖKTA ORD

Under "Sökta ord" kan redaktören ställa in ett antal kriterier för att plockar fram data ut sökloggen. De inställningar som kan göras är följande: möjlighet att välja hur många sökningar som skall visas, möjlighet att se hela fraser eller separata ord, möjlighet att välja att se totalt antal sökningar oberoende av tid, att se sökningar som gjorts under den senaste veckan, den senaste månaden eller att helt fritt välja startdatum och slutdatum. Som standard hämtas totalt antal sökta fraser oberoende av tid. Redaktören kan sedan ändra inställningar.

Resultatet presenteras i en lista med fyra kolumner. Den första kolumnen är rank som bestäms av hur många sökningar som gjorts på resp. ord/fras. Nästa kolumn är ordet/frasen, därefter kommer antal sökningar som gjorts på ett och samma ord/fras och sist presenteras om ordet/frasen finns i synonym- eller rättstavningsordlistan. Alla kolumner utom den sista är sorteringsbara. Genom att klicka på kolumnens rubrik sorteras listan efter innehållet i den kolumnen. Genom att klicka flera gånger görs en omvänd sortering.

Varje ord/fras i den andra kolumnen är länkad direkt till en sökning av ordet/frasen där sökresultatet presenteras tillsammans med hur många klick som gjorts på respektive sökresultat.

Även i den sista kolumnen finns funktionalitet inbyggd. Olika symboler visar om ordet/frasen finns med i någon av ordlistorna. En synonym markeras med  och en felstavning med . När muspekaren förs över respektive symbol visas information om vilka synonymer som finns inmatade till ordet respektive information om vilket ord som det är felstavning av. De ord som inte finns i någon av ordlistorna symboliseras av + . Alla tre symboler är klickbara. Genom att klicka på respektive symbol kan redaktören redigera synonymerna till ordet, redigera felstavningar till ordet samt markera ordet som en synonym eller felstavning.

SÖKNING UTAN TRÄFFAR



Under "Sökning utan träffar" kan redaktören se en lista med sökfraser som besökare använt som inte ger några sökträffar. Listan uppdateras varje gång redaktören går in på menyalternativet. Det kan därför ta lite längre tid att öppna denna sida.


Resultatet visas i en lista med tre kolumner. Den första kolumnen är ordet, därefter kommer antalet sökningar som gjorts på det ordet och sist presenteras om ordet finns i synonym- eller rättstavningsordlistan. Funktionaliteten för denna kolumn är densamma som för samma kolumn under "Sökta ord".


SYNONYMORDLISTAN

Synonymordlistan innehåller ord som ska visas som alternativa sökord när en besökare söker på något av orden i ordlistan. En synonymgrupp kan bestå av två eller flera ord.

Användaren skapar en ny synonymgrupp genom att mata in två ord som har liknande betydelse i textrutorna längst upp på sidan och klicka på "Lägg till".

Nedanför visas rättstavningsordlistan i en lista med två kolumner. Den första kolumnen innehåller ett ord. Nästa kolumn innehåller felaktiga stavningar/benämningar till ordet. Till varje rad finns en redigera knapp - , och en radera knapp - .



Genom att klicka på  öppnas ett nytt fönster där användaren kan lägga till och ta bort felaktiga felstavningar/benämningar från ordet.


Ett ord raderas från ordlistan genom att klicka på . Då ett ord raderas även alla dess tillhörande felstavningar/benämningar.


RÄTTSTAVNINGSORDLISTAN

Rättstavningsordlistan innehåller ord som ska visas som alternativt sökord då en besökare använder ett sökord som är felaktigt. Varje ord kan ha flera felaktiga stavningar/benämningar kopplade till sig.

Användaren skapar ett nytt ord med tillhörande felaktig stavning/benämning genom att mata in två ord i textrutorna längst upp på sidan och sedan klicka på Lägg till.

Nedanför visas rättstavningsordlistan i en lista med två kolumner. Den första kolumnen innehåller ett ord. Nästa kolumn innehåller felaktiga stavningar/benämningar till ordet. Till varje rad finns en redigera knapp - , och en radera knapp - .

Genom att klicka på  öppnas ett nytt fönster där användaren kan lägga till och ta bort felaktiga felstavningar/benämningar från ordet.

Ett ord raderas från ordlistan genom att klicka på . Då ett ord raderas även alla dess tillhörande felstavningar/benämningar.

INSTÄLLNINGAR

Språkstöd

Under inställningar kan användaren ställa in om språkstödet skall vara aktivt. Språkstöd används då webbsidan är byggd i flera språk och har en söksida för varje språk. När språkstödet aktivt finns en synonymordlista, en rättstavningsordlista och en söklogg för varje språk.

Intern Ip-adress

Här matas början på den ip-adress som skall betraktas som intern. (Exempelvis 192.168.0) Sökningar gjorda från adresser som börjar på den interna ip-adressen kan väljas att sorteras bort från sökloggen. Denna inställning är avancerad och behöver inte användas.

Visa sökningar gjorda...

Denna inställning styr om externa och/eller interna sökningar skall visas i sökloggen. Denna inställning är avancerad och behöver inte användas.

Bortsorterade tecken

De tecken som anges här sorteras bort då en besökare gör en sökning som innehåller något av dessa tecken. Tecknet ersätts istället med ett mellanslag.

Antal dagar att spara sökloggen

Här anges hur länge sökloggen ska sparas. Sökningar som gjorts tidigare rensas bort. Rensningen sker genom ett Schemalagt Jobb i EPIServers admin-gränssnitt. Man kan där välja att köra jobbet manuellt eller att schemalägga det. Rekommenderat är att schemalägga det att köra en gång per dag.

ADMINISTRATIONS LÄGE I EPISERVER

RENSA SÖKLOGGEN

I Adminläget till EPIserver finns ett schemalagt jobb för att rensa sökloggen. Varje gång jobbet körs rensas gamla sökningar från loggen. Under AvanSearch (i redigeraläge) → Inställningar, ställs in hur länge (antal dagar) en loggad sökning skall sparas.

FUNKTIONALITET PÅ EN SÖKSIDA

INSTÄLLNINGAR I REDIGERALÄGE

Under fliken Gemensamma -> "Sök", i redigera läge finns möjlighet att anpassa sökmotorns funktionalitet och utseende.

Startsida för sökning

Anger den nod som sökningen skall utgå från. Enbart sidor som ligger under denna nod kommer med i sökresultatet.

Kategori 1-10

Anger de kategorier som finns under avancerad sökning samt de kategorier som sökresultatet delas upp i när användaren väljer att sortera sökresultatet på kategori. Om inga kategorier har matats in till inställningen visas inte detta fält alls under avancerad sökning. Alla sökträffar kommer då att hamna under rubrikern "Övrigt" när en besökare sorterar sökresultatet på kategori.

Sökbara sidtyper

Anger vilka sidtyper som får ingå i sökningen. Sidtyperna matas in med id-nummer och separeras med kommatecken. Exempelvis: 25, 26, 30

Om fältet lämnas tomt sker sökning på alla sidtyper.

Visa "Publicerad" under avancerad sökning

Inställning för att visa/dölja Publicerad under avancerad sökning.

Visa "Andra sökalternativ" under avancerad sökning

Inställning för att visa/dölja "Andra sökalternativ" under avancerad sökning. Andra sökalternativ är i sin tur en editor där valfri text matas in.

Exkludera nod 1-5

Anger noder som inte ska komma med i sökresultatet.

Katalog för Index Server, Scope för IndexServer sökning

Avancerade alternativ för filsökning. Ange den katalogen som angivits i index server.

RESULTAT

RESULTAT

Avantime ville ha hjälp med att paketera och vidareutveckla den produkt de använde som sökmotorlösning till sina kunder. De hade ett antal önskemål på den färdiga produkten. Alla dessa önskemål uppfylldes och Avantime har levererat den färdiga produkten till ett antal av sina kunder.

DISKUSSION

Sedan detta arbete inleddes under 2005 har ett antal kommersiella sökmotorprodukter kommit in på marknaden. Bland de främsta är Google Search Appliance (GSA), SiteSeeker samt Microsofts sökmotor som finns inbyggd i Sharepoint-plattformen. Det har blivit allt vanligare att företag investerar i dessa produkter istället för att använda den enklare sökmotor som finns inbyggd i EPiServer.

Dessa produkter har många av de funktioner som Avantime efterfrågade som tillägg till EPiServers inbyggda sökmotorn. Den avgörande skillnaden mellan EPiServers inbyggda sökmotor och de övriga produkterna är att EPiServer indexerar innehållet i databasen direkt, medan de övriga produkterna istället crawler hemsidorna och indexerar innehållet som de hittar via länkar. En stor fördel med den metoden är att relevansen i sökresultatet blir mycket bättre. De sidor som finns högre upp i menystrukturen rankas högre än de som ligger gömda längre ner i menyn, medan EPiServer inte gör någon skillnad på det. Även hur innehållet på sidan ser ut, tex rubriker, samt hur länken till en viss sida ser ut påverkar relevansen.

Genom att använda andra sökmotorer än EPiServers inbyggda får kunden automatiskt en sökmotor som ger bättre relevans i sökresultatet utan att behöva lägga någon tid eller tankekraft på att styra det själva.

En annan fördel med de kommersiella produkterna är att de har inbyggt avancerat stöd för rättstavning och synonymhantering. Både SiteSeeker och GSA har väldigt bra stöd för det. GSA främst för engelska, men SiteSeeker har mycket bra stöd även för svenska. Microsofts produkt är lite sämre i det avseendet.

När det gäller loggning av sökningar så har SiteSeeker och Google bra stöd och även bra gränssnitt för att analysera loggarna. Microsofts verktyg är mer begränsat i det avseendet.

SLUTSATS

Produkten som togs fram i detta arbete uppfyllde önskemålen från Avantime väl. Dock har det under senare åren kommit ett antal andra kommersiella produkter som har visat sig vara svåra att konkurrera med. Många funktioner som tagits fram i detta arbete kan appliceras även med någon av de andra produkterna som bas istället för EPiServers sökmotor.

REFERENSER

1. Microsoft Visual C#.NET steg för steg
ISBN 9163607166
2. Microsoft Visual C++ .NET steg för steg
ISBN 9163607158
3. Microsoft ADO .NET steg för steg
ISBN 9163607190
4. Developing Solutions with EPiServer
ISBN 91-631-5351-3
<http://world.episerver.com/Documentation/Items/Developing-Solution-with-EPiServer/>
5. EPiServer World
<http://world.episerver.com/>
6. EPiServer
<http://www.episerver.com>
7. SiteSeeker
<http://www.siteseeker.se>
8. Google Search Appliance (GSA)
<http://www.google.se/enterprise/search/gsa.html>