

Utvärdering

Findwise | Solr i SiteVision

Findwise Göteborg
2011-10-26

Innehållsförteckning

1	Bakgrund.....	2
2	Solr i SiteVision	2
2.1	Funktionalitet från Solr i SiteVision	2
2.2	Begränsningar från SiteVisions Solr-paketering	2
2.3	Föreslagen strategi kring SiteVision och Findability	3

1 Bakgrund

Mot bakgrund av att vi har ett flertal kunder som använder SiteVision och att SiteVision 3 inkluderar sökfunktionalitet i form av inbyggd Solr så har vi undersökt och sett på möjligheten att använda den inbyggda Solr för att bygga en Findability-lösning som motsvarar de krav på funktionalitet och källintegration som Findwise kunder efterfrågar. Slutsatserna utifrån den information vi besitter idag om Solr i SiteVision 3 är dokumenterade nedan.

2 Solr i SiteVision

SiteVision har gjort en satsning på att förbättra söket i sin produkt och inkluderar därför Solr med produkten SiteVision 3. Paketeringen mellan SiteVisions grundfunktionalitet och söket från Solr har skett genom att göra Solr "embedded", eller inbyggd, i SiteVisionen produkten och därför nära knuten till den funktionalitet som SiteVision erbjuder.

2.1 Funktionalitet från Solr i SiteVision

SiteVisions inbyggda Solr har fått en funktionalitet som är integrerad och ihopkopplad och som ger standardfunktionalitet kring indexering av innehåll som ska göras sökbart (från dokument som publiceras i SiteVision), hantering av sökfunktionalitet kring facetter (filtrering), paginering och fält och standardrelevans för söket. Ett par av de olika sökfunktionerna (fyra stycken), hanteras av s.k. moduler/portlets som kommunicerar med Solr och ger gränssnittsvyn av resultatet från sökmotorn.

2.2 Begränsningar från SiteVisions Solr-paketering

Det finns ett antal begränsningar i hur Solr har paketerats med SiteVision i avseendet att återanvända Solr-lösningen för t ex ytterligare källor eller utökning av funktionalitet för söket. Dessa begränsningar återfinns i just paketeringen, att Solr lagts in som "embedded" utan kommunikationsmöjligheter mot andra gränssnitt eller applikationer (dvs via http eller API). I Solrs standardutformning, och med en best-practice uppsättning av Solr, så körs Solr som en s.k. webbapplikation som har kommunikationsmöjligheter via http och standard API:er då lätt kan utökas med komponenter som utökar och förbättrar arkitekturen.

Med en inpaketerad Solr likt den uppsättning SiteVision valt att göra så kan Solr därmed inte anpassas, utökas eller kommuniceras med utanför de API:er som SiteVision utvecklat och valt ut – vilket föranleder att indexering av andra sorters källor (annat än SiteVision) inte blir möjligt samtidigt som ytterligare sök-komponenter eller moduler för t ex dokumentprocessering inte kan integreras med den Solr-lösningen. Avsaknaden av komponent/funktionalitet för dokumentprocessering i SiteVisions lösning ger bl a begränsningar i att arbeta med att ta fram en bra relevans, arbeta med meta-data för facetter samt normalisera och berika innehållet. Detta är inte nödvändigtvis ett problem för innehållet som kommer från SiteVision och blir sökbart, men för externa och andra källor är dokumentprocesseringen nödvändig för att jobba med att normalisera innehåll, ta ut lämplig metadata till facetter samt arbeta effektivt med relevansförbättringar.

2.3 Föreslagen strategi kring SiteVision och Findability

I dagsläget och baserat på den information vi besitter om Solr i SiteVision 3 i dagsläget föreslår vi att våra kunder kompletterar SiteVision med en fristående Solr-instans om de vill integrera ytterligare källor, processera dokument eller på andra sätt vill anpassa sökfunktionaliteten eller arkitekturen.

Den fristående Solr-instansen som då kan sättas upp kan i sin tur kommunicera med motorn i SiteVision genom en s.k. "proxy-modul" som SiteVision utvecklat. Denna modul gör det möjligt att kommunicera med externa system, likt en fristående Solr applikation. På så vis kan det innehåll som indexerats i den fristående Solr-lösningen även göras tillgängligt för sök direkt i gränssnittet i SiteVision. För att skapa integrationen krävs att proxy-modulen används.

Den nära kopplingen mellan en fristående Solr med SiteVision skulle fortsatt kunna ge ett enhetligt sök och sökgränssnitt, samtidigt som specifika sökfrågor kan ställas mot båda Solr-instanserna samtidigt. SiteVisions inbyggda Solr ansvarar för sök i det publicerade SiteVision innehållet, och den fristående Solr-instansen för alla andra källor och filer som ska göras sökbara – separata eller helt gemensamma sökgränssnitt/sidor och sökdialoger i gränssnittet i SiteVision kan skapas på så sätt.

Specifika sökinställningar kring boostning, relevans, synonymer m.m. kan synkroniseras mellan de båda Solr-instanserna så att de hanterar även detta på ett enhetligt sätt. Vi har inte sett närmare på uppsättningen av sökstatistik i SiteVisions lösning och hur just den specifika delen skulle kunna synkroniseras med sökstatistik från fristående Solr, om behovet av att denna funktion integreras behöver det undersökas närmare.