

Sök i SiteVision 3

Magnus Lövgren
Niclas Hedlund



Agenda

- Bakgrund
- Sök i SiteVision
- Paketering
- Demonstration
- Indexering av externt material
- Uppgradering från 2.6 / 2.7 (bonusmaterial)

Bakgrund

- Sökfunktionen har funnits med sen 2003
 - Apache Lucene
 - Mycket specialskriven kod
- Begränsade inställningsmöjligheter
 - Ej konfigurerbart indexeringssteg
 - Komplicerad sökmodul
- Lanseringen av "Nya sök" i SiteVision 3
 - Apache Solr
 - Fokus på kundnytta

Sök i SiteVision

- Baserad på Apache Solr
 - Kraftfull, beprövad
- Dataextrahering med Apache Tika
 - Många filformat, metadata i filer
- Helt integrerad i SiteVision
 - Administration via GUI
 - Konfiguration
 - Övervakning

Sök i SiteVision (forts.)

- Loggning och statistik
- "Temporära index"
- Exponering via SiteVisions API
- Lastdelning

Moduler i SiteVision

- Sökning efter information
 - Sök, Sök (facetterad), Sökruta
- Navigation
 - A-Ö lista, Taggmoln
- Speciallösningar
 - Söklistning, skriptmodul
- Redaktionell användning
 - Blogg

Paketering



| | CMS | Portal | EP |
|--|------------|---------------|-----------|
| Sökordsförslag - "Auto complete" | ✓ | ✓ | ✓ |
| Direktsökning - "Live sök" | ✓ | ✓ | ✓ |
| Stavningsförslag - "Menade du?" | ✓ | ✓ | ✓ |
| Sökning på ordstammar - "bil", "bilar" | ✓ | ✓ | ✓ |
| Sökordsmarkering | ✓ | ✓ | ✓ |
| Sortering av sökträffar | ✓ | ✓ | ✓ |
| Behörighetsstyrda sökresultat | ✓ | ✓ | ✓ |
| Alltid aktuellt sökresultat | ✓ | ✓ | ✓ |
| Omindexering | ✓ | ✓ | ✓ |
| Sitemaps | | ✓ | ✓ |
| Lastdelning vid klustring | | | ✓ |

Sökpaket 1

| | CMS | Portal | EP |
|-------------------------------------|------------|---------------|-----------|
| Förfining av sökresultat – facetter | | T | T |
| Sökning på synonymer | | T | T |
| Prioriterade sökträffar | | T | T |
| Loggning av sökningar | | T | T |
| Statistik över sökningar | | T | T |
| Integration av externa sökmotorer | | T | T |

Sökpaket 2

| | CMS | Portal | EP |
|--------------------------------------|------------|---------------|-----------|
| Sökning på flera webbplatser | | T | T |
| Specialanpassade sökindex | | T | T |
| Språkhantering i sökindex | | T | T |
| Anpassning av funktionen "Menade du" | | T | T |
| Anpassning av stoppord | | T | T |
| Indexering av externt material | | T | T |

Demonstration

Indexering av externt material

- Utvecklingsarbete pågår
- Aktuella utmaningar
 - Aggregering
 - "Relevanta delar" av resurser
 - Klustringskoncept
- Egna sökindex
- Externa SiteVision webbplatser
 - Notifiering vid förändring
 - Behörigheter

Egna sökindex

- Öppnar för kundspecifika index
 - Konfiguration via GUI
 - Uttökade inställningsmöjligheter
- Prestandavinster
- Hantering av webbplatser med olika språk
- Sökning på flera webbplatser samtidigt
- Val av index i respektive "sökmodul"

Demonstration

- Sökning på material från extern webbplats
- Behörighetsstyrning av sökträffar

Uppgradering från 2.6 / 2.7

Sökskript och egna portlets

- Borttagna klasser Lucene
 - Hits, Hitlterator, Hit
- Sökskript mot interna klasser
 - SearchEngine.search(String aQuery)
 - SearchEngine.search(Query aQuery)
- Använd SiteVisions publika API!
 - SearchUtil.search(...)

Exempel på söskript

~~// Sökning mot SiteVisions interna klasser~~

```
var hits = CommonServices.getSearchEngine().search("hello"); // Lucene Hits
var hitsIter = hits.iterator(); // Lucene HitIterator

while(hitsIter.hasNext())
{
    var hit = hitsIter.next(); // Lucene Hit
    var doc = hit.getDocument(); // Lucene Document
    out.println("Title: " + doc.get("title"));
}
}
```

// Sökning via det publika API:et

```
var searchUtil = request.getAttribute("sitevision.utils").getSearchUtil();
var result = searchUtil.search("hello");
var hitsIter = searchResult.getHits();

while(hitsIter.hasNext())
{
    var searchHit = hitsIter.next();
    out.println("Title: " + searchHit.getField("title"));
}
}
```


Tillgängliga fält i sökindexet

- Ej analyserade fält lagras utan suffix
 - "title" tidigare "title.untok"
- Analyserade fält lagras med suffix
 - "title.analyzed" tidigare "title"
- Metadatafält (DC.Title) med prefix
 - "metadata.DC.Title"
 - "metadata.analyzed.DC.Title"