

## ARCHITECTUUR GRIDWALKER ZOEKMACHINE

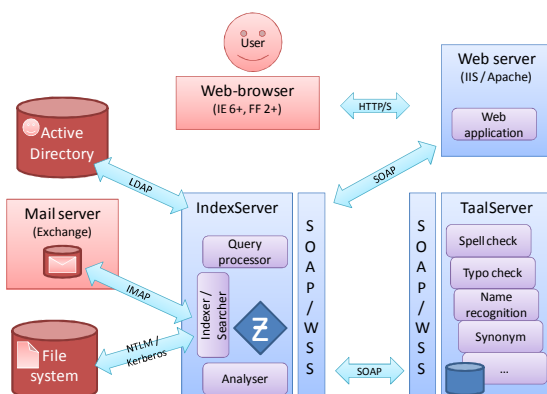
### OPEN ARCHITECTUUR, STANDAARDEN

Openheid is het belangrijkste uitgangspunt van GridWalker Search. Een open architectuur, uitgaande van algemene standaarden.

De belangrijkste onderdelen van GridWalker Search zijn:

- IndexServer voor indexeren, zoeken en taal-technologische verrijking
- GridLine Taalserver modules voor het Nederlands
- Interoperability middels Webservices SOAP/WSS-interface
- Standaard dataconnectoren voor de meeste databronnen
- Web-applicatie voor gebruikersinteractie

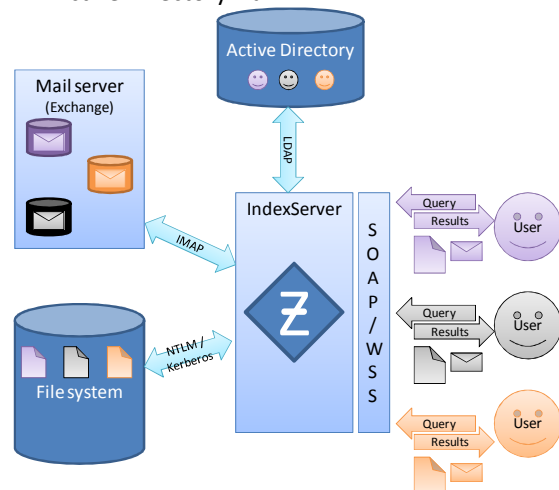
In vijandige omgevingen wordt de integriteit en de vertrouwelijkheid van de berichtuitwisseling met de webservice gegarandeerd middels het standaard communicatieprotocol Web Services Security (*WSS-protocol*). Het WSS-protocol maakt gebruik van XML Signature, XML Encryption, Kerberos, en certificaatformaten zoals X.509.



### INTEROPERABILITEIT

GridWalker Search integreert standaard met o.a.

- Microsoft Exchange Server via IMAP
- Novell GroupWise via IMAP
- IBM Lotus Domino via IMAP
- Active Directory via LDAP



Voor de toegangsrechten tot bestanden en e-mails maakt GridWalker Search standaard gebruik van Active Directory. GridWalker Search ondersteunt hiervoor zowel het Windows specifieke *NTLM* authenticatieprotocol als het veiligere protocol *Kerberos*. Toegang tot externe websites is al naar behoefte op applicatieniveau of individueel gebruikersniveau te specificeren.

De Taalserver en IndexServer zijn volledig in Java ontwikkelde componenten. Beide draaien op alle platforms, waaronder *Windows NT/2000/2003* en *Unix/Linux*. Vanzelfsprekend draait GridWalker standaard ook onder *VMWare*.

De Web-applicatie draait onder zowel Microsoft *Windows Internet Information Service (IIS)* als *Apache Tomcat*.

- Voor IIS is een .Net applicatie beschikbaar
- Voor Tomcat is een Ruby-on-Rails applicatie beschikbaar

## WEB-BROWSERS

De webapplicatie maakt gebruik van JavaScript ten behoeve van het verhogen van het gebruiksgemak en verbeteren van de gebruikservaring.

Onder de ondersteunde web-browsers vallen:

- Microsoft Internet Explorer (IE6, IE7, IE8)
- Mozilla Firefox (FF2, FF3)
- Nieuwere versies van Opera, Safari en Google Chrome.

## STANDAARD DATACONNECTOREN

De IndexServer biedt vele standaard dataconnectoren om tekstuele data te extraheren uit verschillende soorten databronnen. Naast deze standaard dataconnectoren kunnen specifieke connectoren eenvoudig worden geïntegreerd of op verzoek worden ontwikkeld.

Dankzij de standaard dataconnectoren kunnen de meeste databronnen worden ontsloten. De onderstaande protocollen en formaten worden ondersteund.

### Protocollen

- IMAP t.b.v. mailboxen (Exchange, GroupWise, Lotus Domino)
- SRU t.b.v. bibliotheeksystemen
- HTTP/HTTPS t.b.v. wiki's en externe websites
- ODBC/JDBC t.b.v. databases

### Formaten

- Microsoft Office 2000/2003 (MS-Word, MS-PowerPoint, MS-Excel)
- Microsoft Office 2007 (MS-Word, MS-PowerPoint, MS-Excel)
- Microsoft Visio
- Microsoft Outlook Personal Folder File (PST)
- Open Office (Writer, Calc, Impress)
- Adobe Portable Document Format (PDF)
- HTML / Tekst (ASCII, UTF-8, UTF-16)
- E-mail / MIME (ELM, MSG)
- Rich Text Format

## AANVULLENDE EISEN AAN OMGEVING

De interfaces voor de webservices van de IndexServer en de Taalserver staan toe dat beide servers in eigen omgevingen draaien. Wel is toegang via http (poort 80) of https (poort 443) tot de servers noodzakelijk. Het spreekt voor zich dat de algehele performance verbetert indien beide servers wel bij elkaar staan.

De Web-applicatie vereist het .Net-framework (versie 3.5). De IndexServer en TaalServer draaien onder Java Run-time environment (JRE versie 1.6).

Noch de IndexServer, noch de TaalServer stelt hoge eisen aan de hardware. De performance van de IndexServer is te allen tijde eenvoudig te verhogen door de RAM-capaciteit uit te breiden. Voor de index van IndexServer zelf, dient voldoende opslagruimte beschikbaar te zijn. Afhankelijk van de soorten en hoeveelheid taalkundige verrijking van de tekstuele data, kan de totale index tot twee maal de omvang van de geïndexeerde data groeien.

*GridLine – Grip op informatie*  
Keizersgracht 520  
1017 EK Amsterdam  
T: 020 616 2050  
F: 020 616 2051  
E: [gridwalker@gridline.nl](mailto:gridwalker@gridline.nl)  
W: [www.gridline.nl](http://www.gridline.nl)