

Federated Search Engines, een studiedag die smaakt naar meer...

Brussel, Erasmushogeschool, 5 mei 2009

Er is een groot aanbod van federated search engines op de markt. Waar moeten bibliotheken op letten bij de aankoop van deze zoekmachines, wat zijn de voor- en de nadelen? Tijdens de infosessie die de sectie Schoolbibliotheken van de VVBAD op 5 mei organiseerde in de Hogeschool Brussel, gingen leveranciers en gebruikers dieper in op de mogelijkheden van deze nieuwe tool.

Een van de kenmerken van het huidige informatielandschap is de uitbundige groei van het aantal websites en online databases. Voor wie op zoek is naar relevante, contextspecifieke en op maat gemaakte informatie, is het vaak een hele opgave om de weg te vinden in het doolhof van alle beschikbare bronnen. Met federated search engines (FSE) kunnen gebruikers in één duidelijke en samenhangende zoekomgeving diverse informatiebronnen raadplegen.

Naar goede gewoonte verwelkomde **Sissi Loostermans**, voorzitter sectie Schoolbibliotheken, de aanwezigen, stelde de sprekers voor en formuleerde achteraf de conclusies van deze interessante studiedag.

Luk Laermans, departementshoofd van de Erasmushogeschool Brussel, benadrukte dat informatievaardigheden steeds belangrijker worden in de huidige context van onderwijs, onderzoek, diensten aan derden, levenslang leren, afstandsonderwijs én besparingen. Daarmee zette hij ineens de rol van de bibliotheek als coördinatiecentrum van het digitale informatieaanbod, waar informatieprofessionals als coach optreden bij het zoeken en verwerken van informatie, op de voorgrond.

Paul Nieuwenhuysen maakte op de hem eigen, aangename en vooral boeiende manier een aantal kritische bedenkingen bij de FSE. Hij wees erop dat ook websites voor hotelboekingen, luchtvaartmaatschappijen en musea de techniek van FSE gebruiken in hun toepassingen. Hij somde eerst een aantal voordelen op:

- FSE kunnen de gebruikers helpen de juiste bronnen te selecteren;
- mensen moeten minder 'informatievaardig' zijn;
- er is één user interface, één syntax en een consistente manier om zoekresultaten te tonen;
- er is tijdswinst, omdat niet alle databanken afzonderlijk moeten worden doorzocht;
- niet gekende databanken worden ook doorzocht;
- bij interdisciplinariteit wordt vaak relevante informatie gevonden;
- verder verwerken van de resultaten is mogelijk;
- interessante links worden aangeboden;
- eventuele full text beschikbaarheid wordt getoond.

Kortom, het lijkt alsof je ineens hebt gevonden waar je altijd naar gezocht hebt: in een simpele zoekomgeving tegelijkertijd diverse informatiebronnen raadplegen, terwijl de resultaten op een eenvormige manier getoond worden.

Maar er zijn ook een aantal moeilijkheden, uitdagingen, problemen en beperkingen. In de eerste plaats is er de restrictie op de zoekmogelijkheden: de structuur, de kwaliteit en de mogelijkheden van de doorzochte databanken bepalen de mogelijkheden van de zoekmachines. Bovendien:

- wanneer is hetzelfde resultaat ook echt hetzelfde (probleem van de duplicatie)?
- wat betekent sorteren op relevantie in deze context?
- diepgaand zoeken is niet mogelijk;
- irrelevante informatie geeft tijdsverlies;
- gebruik van verschillende protocollen in datacommunicatie;
- verschillende manieren van werken (booleaans, truncatie);
- vaak lange antwoordtijden;
- gebruikers denken vaak dat ze alles doorzocht hebben, maar misschien is dat niet echt zo;
- gebruikers kennen hierdoor misschien de relevante databanken niet meer;
- soms zitten niet alle databanken in een zoekmachine, omdat het om commerciële redenen niet mag;
- evaluatie van de kwaliteit van de gevonden informatie wordt moeilijk, omdat gebruikers niet meer weten waar de informatie vandaan komt.

Dus blijft de kwaliteit van FSE nog steeds een continue uitdaging voor de softwareontwikkelaars én voor de bibliotheken die een dergelijke tool gebruiken.

Vervolgens stelden **Rene Dales** en **Ron Burns** EBSCOhost Integrated Search (EHIS) voor. Deze tool werkt met een 'tiering'systeem van de doorzochte databanken. Hiermee kan bepaald worden welke databanken in welke vakgebieden het eerst doorzocht worden op basis van verschillende criteria (relevantie, zoektijd).

Als laatste item vóór de middagpauze demonstreerden **Stefaan Renard** (VUB) en **Filip Trensou** (Infor) V Spaces. Filip Trensou ging dieper in op V Space Arachne, een virtueel netwerk van openbare bibliotheken in de provincie Antwerpen, met een portaalomgeving en een enkel toegangspunt tot verschillende bibliotheekcatalogi, databanken en websites. In Croydon College wordt V Spaces geïntegreerd in de Aquabrowser interface.

Stefaan Renard liet zien hoe V Spaces in de bibliotheekwebsite van de VUB is geïmplementeerd.

Na de lunch volgde nog een goed gevuld middagprogramma met vier presentaties. De eerste ging over Elin, het buitenbeentje op deze studiedag. Reden: het is geen FSE. Gebruikers zoeken er slechts in één database, die is samengesteld met metadata aangeleverd door uitgevers, DOAJ en repositories als Cogprints of ArXiv.org. Lund University Libraries zorgt voor de externe hosting. De bibliotheek zorgt steeds zelf voor het beheer van de metadata van tijdschriften, artikelen, e-books, databanken en subject gateways. Door een klik op een referentie in Elin kun je doorklikken naar de fulltext. Via Elin worden al ongeveer 27,5 miljoen

artikelen, 32.000 tijdschriften, 27.000 e-books en 558 databanken ontsloten. Als onderwerpsontsluiting werd gekozen voor LCC (Library of Congress Classification). Het voordeel van Elin ten opzichte van FSE is de tijdsduur, volgens **Inge Van Nieuwerburgh** van UGent. Elin werkt namelijk merkbaar sneller dan de klassieke FSE. De resultaten worden in één overzichtelijke resultatenpagina getoond en de resultaten zijn eenvormiger. Ter afsluiting kregen we een demo van Elin@Gent, een portaal voor louter elektronische artikelen. Studenten krijgen de boodschap dat er geen bibliografische data in beschikbaar zijn. Elin werd op UGent ook geïntegreerd in de bibliotheekcatalogus MeerCat.

Veerle Kerstens demonstreerde LibriSource Plus, de FSE die K.U.Leuven gebruikt als portaal voor haar elektronische bronnen, databanken, e-tijdschriften en e-books. Simultaan kunnen ook catalogi en internetbronnen worden doorzocht. De authenticatie is niet verplicht op de campus. Off campus toegang is mogelijk door inzet van Shibboleth-technologie of via IP-herkenning. In de databank kun je zoeken op voorge-definieerde sets van databanken of zelfsamengestelde sets. Tijdens een zoekactie worden dubbele records wel herkend, maar niet uitgefilterd. Voor de koppeling naar de fulltext wordt onderliggende SFX-technologie gebruikt, onder de naam Librilink.

Het resultaat van een zoekactie kan in LibriSource, maar ook in de native (oorspronkelijke) database worden getoond. De zoekactie kan bewaard worden en het is mogelijk om search alerts aan te maken, om zo met regelmaat en automatisch zoekresultaten te ontvangen in je mailbox. Vanuit de eigen bibliotheekcatalogus of de elektronische leeromgeving kan er gekoppeld worden naar Librisource Plus.

Jo Rademakers stelde het nieuwe project Primo voor. Primo is geen klassieke FSE, maar een centrale index, vergelijkbaar met Elin, van waaruit wordt gelinkt naar tal van bronnen zoals Aleph (catalogus), IAS (institutionele repository), Toledo (de elektronische leeromgeving) en Librilink, maar ook externe databanken en het web. De bedoeling is om aan de gebruiker de informatie veel sneller aan te leveren, op een centrale plaats en in een herkenbare omgeving. Een obstakel is dat uitgever dit niet altijd toestaan. Soms is dit ook technisch niet mogelijk, omwille van de omvang van de data. Dan valt men terug op Librisource. Primo gaat wellicht in september 2009 officieel van start. Tot slot deed Jo Rademakers een oproep om gezamenlijk en op een centrale plaats een index aan te leggen en te onderhouden voor bijvoorbeeld de elektron databanken. De werklust kan zo onder meerdere partners worden verdeeld.

Swets werd vertegenwoordigd door **Olivier Van Kerkhove**. Hun FSE heet SwetsWise Searcher. De interface is compleet aanpasbaar naar de wensen van de gebruiker. Het product is pas opgeleverd in Nederland. In SWS kan de gebruiker de zoekbronnen groeperen in pakketten naar keuze. Deduplicatie van de resultaten is mogelijk. Voor het doorlinken naar fulltext wordt Swetswise Linker gebruikt. De resultaten

worden overzichtelijk getoond in een wolk met geassocieerde woorden, waaruit de gebruiker kan kiezen.

Swets kan nu al vlot koppelen naar 5500 databanken. De voorwaarde om goed te koppelen is dat de databank met XML, HTML of Z39.50 doorzoekbaar is. Ook andere soorten bronnen, zoals catalogi zijn doorzoekbaar, indien ze vrij beschikbaar zijn op internet. Swets belooft een snelle implementatietijd van ongeveer een maand. Wie instapt op SWS krijgt Swets Online Content er als bron gratis bij, waarmee 30 miljoen artikelen ter beschikking komen.

Webfeat is het product dat Hogeschool Gent aankocht. **Sofie Vandroemme** schetste ons de pro's en contra's die hun keuze beïnvloedde. Positief vonden zij de aankoop-prijs, het aantal beschikbare connectoren (databanken waar onmiddellijk mee gekoppeld kan worden), de goede service tijdens de eerste contacten en de snelle implementatietijd. Als nadelig beschouwde ze de limiet van vijftig databanken (waarna een herberekening van de prijs volgt), het feit dat de partner een commercieel bedrijf is en dat sommige databanken wellicht niet kunnen geconnecteerd worden (bijvoorbeeld Kluwer databanken).

Met de management tool kan de beheerder makkelijk zelf databanken toevoegen, op voorwaarde dat de connector al beschikbaar is. De database vendors moeten wel het Webfeat-IP accepteren, dit wil zeggen als geldig IP toelaten, om de databank te doorzoeken. Dit gaf in een aantal gevallen problemen, die soms met vertraging (Swetswise) of soms helemaal niet opgelost geraakten (IEEE, Google Scholar). De databank is sinds maart 2009 in gebruik. Sofie Vandroemme rondde de lezing af met de eerste bevindingen:

- de firma Webfeat levert een zeer goede service;
- het product Webfeat integreert goed met SFX;
- statistieken zijn te verkrijgen per afzonderlijke module, maar tegen een prijskaartje;
- resultaten verschijnen soms zeer traag op het scherm;
- er treden verschillen op tussen de resultaten uit de native databank en Webfeat;
- de zoekmogelijkheden zijn beperkter wanneer de gebruiker in Webfeat zoekt t.o.v. in de native databank;
- het product is vooral interessant voor de 'zoekende student', die zich nog moet oriënteren in een onderwerp, maar minder geschikt voor de onderzoeker.

De presentaties van de informatiedag zijn te vinden op de website van VVBAD, <http://www.vvbad.be/node/4446>.

Dirk Desaeveer, Diensthoofd EMI Plantijn Hogeschool
dirk.desaeveer@plantijn.be

Chris Goetschalckx, Bibliothecaris-coördinator Artesis Hogeschool Antwerpen
christiane.goetschalckx@artesis.be