

# **Review inzet Aquabrowser in het Theasaurus-project**

Dr. Peter B. Sloep  
Fontys Hogescholen  
Lectoraat Educatieve Functies van ICT

## **Opdrachtomschrijving**

Thesaurus is een project dat is uitgevoerd onder leiding van de Fontys Hogeschool Communicatie in samenwerking met de School voor Communicatiemanagement van de Hogeschool Utrecht. Het project is uitgevoerd tussen september 2001 en december 2003. De Stichting SURF en de Digitale Universiteit hebben het project financieel gesteund.

Thesaurus heeft als doel het onderwijs sterk te verbeteren door 'de muren tussen onderwijsinstellingen te slopen'. Daartoe is een kennisnetwerk op gebied van communicatie opgezet, opdat studenten en docenten van verschillende instellingen kennis met elkaar kunnen delen. Zij doen dat door de (impliciete) kennis waarover zij beschikken binnen de kaders van het netwerk te expliciteren en zo voor de andere leden van de gemeenschap te ontsluiten. Ontsluiting van (expliciete) kennis is de ene kant van de medaille, vinden van de kennis die je zoekt is de andere. Een goed functionerend kennisnetwerk staat of valt met het vermogen relevante kennis in een kennisbank te zoeken en te vinden. Na een aanvankelijke vrijage met Ariadne-technologie van ongeveer een half jaar, heeft de projectleiding besloten gebruik te maken van de AquaBrowser om de gewenste functionaliteit voor de kennisbank te verschaffen. Ariadne leverde niet voldoende kwaliteit, meer in het bijzonder (zie bijlage 1, *Van Ariadne naar AquaBrowser*):

- de performance van de Ariadne-servers liet te wensen over
- de helpdesk van Ariadne was vaak niet bereikbaar
- de Ariadne-technologie paste bij nader inzien niet op de eisen die het Thesaurus-project stelt aan zoekfunctionaliteiten
- de AquaBrowser-technologie paste bij nader inzien wel bij die eisen.

De projectleiding beloofde onderstaande expert review met als bedoeling 'een buitenstaander kritisch [te] laten kijken naar de meerwaarde van de AquaBrowser als zoekmachine [...]' (zie bijlage 1, *Van Ariadne naar AquaBrowser*).

Meer specifiek zal ik in dit document de vraag beantwoorden of het besluit de Ariadne-technologie te verlaten ten faveure van de AquaBrowser-technologie verstandig is geweest *in het licht van de eisen die het Thesaurus-project stelt aan de kennisbank*. Dit betekent dat ik overwegingen van server performance, bereikbaarheid van de helpdesk en dergelijke buiten beschouwing laat. Dit soort overwegingen, die de operationele inzet van een technologie betreffen, hangen vooral samen met de manier waarop een technologie geïmplementeerd wordt. En hoewel een goede implementatie een innovatietraject kan doen slagen of mislukken, staat de vraag of een gegeven technologische oplossing past bij de vereisten die men heeft opgesteld, daar los van. Ik zal me daarom beperken tot

- (i) een inventarisatie van de eisen die het project stelt aan de technologie die onder de kennisbank ligt en
- (ii) een analyse van de vraag in hoeverre Ariadne en AquaBrowser aan die eisen kunnen voldoen.

## **Eisen**

Het controlling document, waarvan het eerder genoemde document 'Van Ariadne naar AquaBrowser' als bijlage 3 deel uitmaakt, bevat geen lijst met vereisten. Maar het bevat genoeg materiaal om tot zo'n lijst te komen (de nummers tussen () verwijzen naar paginanummers in het Controlling document). Eisen die aan de kennisbanktechnologie gesteld worden (1 - 6) zijn onderscheiden van eisen die specifiek aan de onderliggende zoektechnologie worden gesteld (7 - 12).

- 1 De technologie moet het *formeren van een kennisbank* mogelijk maken, waarin informatie vanuit het onderwijs en het beroepenveld bijeengebracht en effectief teruggevonden kan worden (4).
- 2 De kennisbank moet niet bij uitstek geschikt zijn voor 'het samenstellen van onderwijsmodules via hergebruik van bestaand onderwijsmateriaal en voor het uitserveren van deze cursussen naar groepen studenten (12). Positief geformuleerd houdt deze eis in dat de technologie het samenstellen van *collecties van documenten* moet ondersteunen.
- 3 Aan de bank moet *heterogene informatie* kunnen worden toegevoegd: bronnen (vakliteratuur),

- beschrijvingen van personen (who is who), hyperlinks (5, 13).
- 4 De kennisbank moet *ondersteuning bij vakinhoudelijke opdrachten* bieden (5).
  - 5 De kennisbank kan studenten helpen zich te *oriënteren op hun scriptieonderwerp* (5), in het bijzonder kan de gebruikte zoektechnologie ook probleemstellingen in scripties opsporen en afzonderlijk presenteren (6).
  - 6 De kennisbank moet *niet* bij uitstek geënt zijn op een model waarin docenten informatie aanleveren en studenten informatie consumeren; de bank moet daarentegen een *egalitair model* van informatieaanlevering en -consumptie mogelijk maken, waarin ook een rol voor buitenstaanders is weggelegd (12).
  - 7 Er moet volstaan kunnen worden met *metadatering* van aan de kennisbank aangeleverde informatie volgens de Dublin Core-specificatie; metadatering volgens de IEEE LOM moet geen vereiste zijn (12, 13).
  - 8 De bij de kennisbank behorende gebruikte zoektechnologie kan worden uitgerust met een *eigen lijst van zoekwoorden* (8).
  - 9 De gebruikte zoektechnologie staat *associatief zoeken* toe, waarbij de zoekende niet precies kan aangeven waarnaar hij of zij op zoek is, daarvan een vage notie heeft, maar al doende tot een steeds scherper omlijnende zoekvraag komt (8, 12).
  - 10 De gebruikte zoektechnologie is *aantrekkelijk* voor studenten, docenten en nieuwe partners (9, 13).
  - 11 De zoektechnologie moet een hoge *recall* (nauwelijks relevante documenten missen - fout-negatieven) hebben zonder veel concessies aan de *precision* (nauwelijks vervuild zijn met irrelevante vondsten - fout-positieven) te hoeven doen (13).
  - 12 De zoektechnologie moet resistent zijn tegen *linguïstische varianten* (typefouten, afwijkende spellingen, termen met een zelfde stam) (13).

De lijst maakt duidelijk dat de eisen die gesteld worden aan de technologie volgens welke de kennisbank wordt opgebouwd los staan van de eisen die gesteld worden aan de technologie volgens welke de bank vervolgens doorzocht wordt. In het Controlling document wordt dit onderscheid niet gemaakt, waardoor de argumentatie voor of tegen Ariadne dan wel AquaBrowser soms niet zuiver is. Hierover later meer.

## **De kandidaattechnologieën**

### ***Ariadne***

Het Ariadne-project (er waren er eigenlijk 2) is gesponsord geweest door Europese sector Telematics for Education and Training van het 4e Kaderprogramma van de Europese Unie. Het project richtte zich op 'the development of tools and methodologies for *producing, managing and reusing computer-based pedagogical elements and telematics supported training curricula*' (zie het onderwerp 'history' op de website van de Ariadne Foundation). Na afloop van deze projecten (2000) heeft men de Ariadne Foundation opgericht met dezelfde doelstellingen als het project. In het kader van Ariadne heeft men diverse instrumenten ontwikkeld, zoals auteursinstrumenten, waarmee leerobjecten kunnen worden aangemaakt, en kerninstrumenten. De kerninstrumenten omvatten bijvoorbeeld het Knowledge Pool System, een repository waarin de leerobjecten worden opgeslagen; het Learner Interface, via welke leerobjecten aan lerenden worden getoond; de Curriculum Editor, die je in staat stelt leerobjecten tot curricula te assembleren. Hoewel Ariadne een van de eerste projecten was die dit soort faciliteiten ging ontwikkelen, zijn er inmiddels veel van dit soort systemen beschikbaar, ook commerciële. De indeling in een auteursomgeving, managementsysteem voor leerinhouden (LCMS) en afspeelomgeving voor leerobjecten (VLE), eventueel gekoppeld met administratieve instrumenten (MLE) is gemeengoed geworden (zie bijvoorbeeld Verstelle et al, 2003a en 2003b). Een in dit verband belangrijk punt is dat Ariadne al vroeg - in 1995 - een specificatie ontwikkeld heeft voor het metadateren van leerobjecten (Duval, 2002). Die specificatie is de uiteindelijke IEEE Learning Object Metadata-specificatie,

kortweg LOM geworden (NEN, 2003). In 1996 is men ook begonnen een indexeringsgereedschap (SILO) te bouwen voor de Knowledge Pool op basis van de metadatabeschrijvingen. Zo werd het ook mogelijk op leerobjecten te zoeken in de Knowledge Pool.

### ***AquaBrowser***

De AquaBrowser-technologie is ontwikkeld door Medialab. De AquaBrowser is een zoekmachine, die je net als andere zoekmachines in staat stelt op basis van ingetypte sleutelwoorden documenten te zoeken in een gegeven documentencollectie. De zoekmachine indexeert de woorden in de documenten. Aan de documenten toegevoegde metadata worden overigens ook geïndexeerd. Die collectie kan documenten bevatten in verschillende formaten, bijvoorbeeld .html, .doc, pdf. De Aquabrowser doorzoekt niet het gehele internet, zoals Google, maar is bedoeld voor het doorzoeken van een relatief beperkte collectie van documenten. Die collectie kan overigens nog steeds heel omvangrijk kan zijn, bijvoorbeeld zo groot als de collectie van de bibliotheek Fryslan of de documenten op de website van de Universiteit Groningen. In haar werking onderscheidt de AquaBrowser zich van gebruikelijke zoekmachines in een aantal opzichten, zowel van zoekmachines die het web doorzoeken als van zoekmachines die relatief beperkte documentencollecties doorzoeken. Het belangrijkste daarvan is misschien wel dat de AquaBrowser associatief ('fuzzy') zoekt. Door gebruik van een aantal taaltechnologische methoden is de AquaBrowser in staat als resultaat van een zoekopdracht niet alleen documenten te tonen waarin de zoektermen letterlijk terugkomen, maar ook documenten met semantisch verwante termen (in dezelfde of een andere taal), synoniemen (antwoord, respons), en termen met dezelfde stam (roos, rozengaur). Welke taaltechnologieën de AquaBrowser precies gebruikt, is waarschijnlijk een bedrijfsgeheim, maar uit de beschikbare documentatie op <<http://www.medialab.nl/index.asp?page=about/technologydetail>> en een achtergrondartikel van medewerkers van Medialab (Anonymous) blijkt dat de AquaBrowser werkt de gebruikelijke stemmers, stoppers en (varianten op) Latent Semantic Indexing (zie Bray, 2003a en Van Brugen et al. in press). De resultaten van een zoekopdracht worden in een resultatenlijst getoond en in een netwerkstructuur. Resultaten worden weergegeven zodat zij de mate van verwantschap met de zoekopdracht weerspiegelen. In de lijst worden de meest verwante resultaten bovenaan gezet, in het netwerk krijgen de meest verwante resultaten de meest prominente positie. De AquaBrowser-zoekmachine kan bovendien werken op basis van een door de gebruiker zelf ontwikkeld ontologie (hiërarchisch systeem van categorieën van termen). Hierdoor kunnen ook heel gerichte zoekopdrachten worden gegeven. Zie ook de website van Medialab < <http://www.medialab.nl/index.asp?page=about/technology>>.

### **Analyse**

Uit de samenvatting van de Ariadne- en AquaBrowser-technologieën blijkt dat zij antwoord geven op geheel verschillende vragen. Ariadne is het antwoord op de vraag naar het faciliteren van hergebruik van onderwijs. Ariadne heeft daarvoor een instrumentarium ontwikkeld dat de gebruiker in staat stelt tot metadatering, opslag van leerobjecten, uitserveren van leerobjecten, etc. Een van die instrumenten is een zoekmachine die werkt op basis van metadata van leerobjecten. AquaBrowser is een (bijzonder soort) zoekmachine die een documentencollectie als uitgangspunt kiest. Dat hier sprake is van antwoorden op verschillende vragen blijkt als we ons realiseren dat Ariadne heel goed de Aquabrowser zou kunnen gebruiken om haar Knowledge Pool te doorzoeken. Daar is zelfs veel voor te zeggen omdat de AquaBrowser kan zoeken met behulp van metadata (dus ook die van de LOM) en gevoed kan worden met een ontologie (die weer ten grondslag kan liggen aan een instantie van het Classifications-element 9 uit de LOM). AquaBrowser biedt waarschijnlijk dus wat de Ariadne-zoekmachine biedt, plus de mogelijkheid tot fuzzy-zoekopdrachten. Bij het beantwoorden van de vraag hoe verstandig de overstap van Ariadne op de AquaBrowser is geweest in the Thesaurus-project is deze constatering van cruciaal belang.

Kijken we nu naar de bovenstaande lijst van eisen en scoren we in hoeverre de Ariadne- en AquaBrowser-technologieën aan elk van de vereisten voldoen (+ voor ja, - voor nee, 0 voor irrelevant)

<i>vereisten</i>	<i>Ariadne</i>	<i>AquaBrowser</i>
<i>kennisbank</i>		
1 formeren ervan	+	0
2 is collectie van allerlei documenten	-	0
3 bevat heterogene informatie	-	0
4 ondersteuning bij opdrachten	+	0
5 oriëntatie op scriptie	+	0
6 egalitair aanlevermodel	-	0
<i>zoekmachine</i>		
7 eenvoudige metadatering	+	+
8 eigen lijst van zoekwoorden	+	+
9 associatief zoeken	-	+
10 aantrekkelijk interface	-	+
11 hoge recall en precision	?	?
12 linguïstische varianten	-	+

De bovenstaande scores behoeve enige toelichting:

- ad 1 - 6 De AquaBrowser biedt geen faciliteiten voor het bouwen en beheren van een kennisbank, al gaat hij natuurlijk wel uit van de beschikbaarheid van zo'n bank. Daarom is op de vereisten 1 tot en met 6 telkens een 0 gescoord.
- ad 2, 3 Omdat Ariadne werkt met leerobjecten scoort het negatief op elk vereiste dat om iets anders dan een leerobject vraagt
- ad 6 In principe zouden ook studenten leerobjecten kunnen aanleveren, maar bij mijn weten is Ariadne daar nu niet op ingericht.
- ad 7 Ariadne werkt weliswaar met een tamelijk uitgebreid toepassingsprofiel van de LOM, maar het staat je uiteraard vrij dat voor je eigen doeleinden aan te passen; je blijft dan LOM-conformant. De mogelijkheid metadata in te voeren en erop te zoeken is belangrijk als aanvulling op full-text searches omdat metadata een beschrijving over een document en dus informatie kunnen toevoegen die via een full-text search niet te ontsluiten valt.
- ad 8 ik interpreteer deze eis als de mogelijkheid eigen ontologieën te gebruiken; de LOM (en dus Ariadne) staan dat toe, de AquaBrowser biedt faciliteiten voor de invoer daarvan.
- ad 10 Deze eis is geïnterpreteerd als een grafische representatie van de zoekresultaten. De Ariadne zoekmachine geeft een lijst van resultaten in de vorm van tekst, AquaBrowser doet dat ook, maar voegt daar een grafisch netwerk aan toe.
- ad 11 Ik beschik in geen van beide gevallen over getallen. Het verkrijgen daarvan is overigens niet eenvoudig (Bray, 2003b).

## **Conclusie**

De Ariadne-technologie is gericht op het ontwikkelen, beheren, zoeken en aanbieden van onderwijsmateriaal (leerobjecten), terwijl Thesaurus behoefte heeft aan een technologie die het inrichten, beheren en doorzoeken faciliteert van kennisbank die allerlei soorten van teksten kan bevatten, waaronder eventueel leerobjecten. Waar voorts in Ariadne het accent ligt op ontwikkelen, beheren en aanbieden, ligt bij Thesaurus de nadruk op zoeken. Oom deze beide redenen is Ariadne een weliswaar mogelijke maar minder voor de hand liggende technologie voor het Thesaurus-project. De AquaBrowser technologie biedt associatieve zoekfunctionaliteit voor grote, maar begrensde documentencollecties. Dat is precies de zoekfunctionaliteit waaraan het Thesaurus-project behoefte zegt te hebben. De AquaBrowser-technologie biedt geen gereedschappen voor het ontwikkelen en beheren van zo'n collectie. In dat opzicht schiet de AquaBrowser dus te kort. Zolang het gaat om relatief beperkte collecties documenten zijn dit soort gereedschappen in ruime mate voor handen in de vorm van tekstverwerkers en door het operating system bijgeleverde bestandssystemen. Voor collecties die uitgebreider worden is een of ander content management systeem op basis van een database

noodzakelijk. Ik kan uit de documentatie niet opmaken waarvoor in het Thesaurus-project is gekozen. Een kritische kanttekening bij de AquaBrowser-technologie is overigens dat een Windows PC als client wordt vereist (Wegman, 2004). Waar gebruik wordt gemaakt van Java en Java Virtual Machines platformonafhankelijkheid mogelijk maken, is het jammer dat het gebruik van het Thesaurus-kennisnetwerk tot Windows-gebruikers beperkt blijft. Dit doet afbreuk aan de de doelstelling van het Thesaurus-project.

Het besluit de Ariadne-technologie te verlaten ten faveure van de AquaBrowser is naar mening per saldo verstandig geweest, met als kanttekening dat de inzet van een content management system voor het beheer van de documentencollectie verstandig zou zijn geweest (even aannemende dat men dat niet gedaan heeft).

#### *Acknowledgements*

Ik wil Medialab Solutions BV bedanken voor het verstrekken van achtergrondinformatie, zowel in mondelinge als schriftelijke vorm. Wat ik in deze review over Medialab schrijf is uiteraard geheel voor mijn eigen verantwoordelijkheid, inclusief mogelijke foutieve beweringen over wat de AquaBrowser al dan niet vermag.

#### **Referenties**

- Anonymous (?) *AquaBrowser Pro version 1.1 AquaBrowser Site Manual*. Medialab Solutions Ariadne Foundation, Website. <<http://www.ariadne-eu.org/>>
- Bray, T. (2003a) *On Search: Backgrounder* <<http://www.tbray.org/ongoing/When/200x/2003/06/15/OnSearch>>
- Bray, T. (2003b) *On Search: Precision and Recall* <<http://www.tbray.org/ongoing/When/200x/2003/06/22/PandR>>
- Duval, E. (2002) *Hergebruik en standaarden* <[http://www.surf.nl/download/Duval\\_Surf\\_Leuven.pdf](http://www.surf.nl/download/Duval_Surf_Leuven.pdf)>
- Fontys Hogeschool Communicatie (2002) Thesaurus ontwikkelt Com2Know, kennisnetwerk voor communicatie. SURF
- NEN (2003) *Conceptnorm voor Leerobject-metadata, Nederlandse versie, 1.0* <<http://www.nen.nl/nl/act/spec/leertech/ieeelom.pdf>>
- Van Bruggen, J., P.B. Sloep, P. van Rosmalen, F. Brouns, H. Vogten, R. Koper, C. Tattersall (aangeboden ter publicatie) *Latent semantic analysis as a tool for learner positioning in learning networks for lifelong learning*
- Verstelle, M., P.B. Sloep & B. de la Parra (2003a) ELO's, DLO's, LMS'en: achtergronden en soorten. In: *ICT in het hoger onderwijs; stand van zaken*, H.Frencken, J.Nedermeijer, A.Pilot. I. ten Dam (red.), IVLOS en ICLON. blz. 85 -98.
- Verstelle, M., B. de la Parra, P.B. Sloep (2003b) De keuze van een elektronische leeromgeving. In: *ICT in het hoger onderwijs; stand van zaken*, H.Frencken, J.Nedermeijer, A.Pilot. I. ten Dam (red.), IVLOS en ICLON. blz. 99 -112.
- Wegman, W (2004) AquaBrowser Library R

#### **URL's**

- Ariadne Foundation <<http://www.ariadne-eu.org/>>  
Medialab <<http://www.medialab.nl/>>

#### **Bijlage 1: Van Ariadne naar AquaBrowser**

(document maakt als bijlag 3 deel uit van het herziene controling document van het Thesaurus-project)