



UvA-DARE (Digital Academic Repository)

Topic driven access to scientific handbooks

Caracciolo, C.

[Link to publication](#)

Citation for published version (APA):

Caracciolo, C. (2008). Topic driven access to scientific handbooks Amsterdam: SIKS

General rights

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <http://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

Samenvatting

“Handbooks” zijn dikke delen, die door verschillend auteurs worden volgeschreven met tientallen pagina’s lange hoofdstukken. Het doel is een omvattend overzicht te geven van een wetenschappelijke vakgebied. Helaas ontbreekt er vaak een duidelijke gestandaardiseerde structuur. In de internationale wetenschappelijke literatuur zijn deze *Handbooks* niet zomaar gelijk te stellen aan het meer encyclopedische Nederlandse handboek. Gebruikers van een wetenschappelijke *Handbooks* kunnen een specifieke informatiebehoefte hebben, wat hen er toe brengt een specifiek onderwerp of zelfs een specifiek aspect daarvan te willen lezen. In het laatste geval wil de lezer liever niet door tientallen pagina’s bladeren, maar zou graag direct naar die tekstdelen van het boek “springen,” die het gezochte aspect behandelen. Dit noemen we gefocusseerde toegang tot een tekst. In andere gevallen kunnen *Handbook* gebruikers vagere informatiebehoeften hebben die meer gerelateerd zijn aan de noodzaak of behoefte om een bredere en algemener overzicht van het vakgebied te krijgen.²

Vanuit de behoefte aan gefocusseerde toegang is het begrijpelijk om te denken aan een overzichtskaart van het vakgebied. Een afzoekbare kaart met zowel de verschillende onderwerpen en hun relaties, alsmede de relevante plaatsen in de tekst waar een en ander behandeld wordt. De gebruiker kan dan door zoeken of door navigatie op een plek in de kaart aankomen en vandaar inzoomen op het onderwerp dat het beste toegesneden is op haar informatiebehoefte. Daarom stellen we ons een kaart voor van het vakgebied dat niet alleen referenties bevat naar de passende locaties in de tekst maar ook de relaties weergeeft tussen de verschillende informatie elementen onderling—we noemen deze *topics*. Onze vraagstelling is nu: Wat kunnen we winnen voor het toegankelijk maken van een *Handbook* als wij een kaart aanbieden die verwijzingen, links, bevat naar het boekwerk? Dit proefschrift heeft tot doel om deze kwestie te beantwoorden. Daartoe splitsen we de vraagstelling in kleinere onderzoeksvragen.

²Op deze plek beschouwen we niet het geval van het volgordelijk doorlezen van het hele boek of een hoofdstuk daaruit.

In Hoofdstuk 1 introduceren wij het onderwerp en definiëren wij de onderzoeksvragen.

In Hoofdstuk 2 ontwikkelen wij onze visie op gefocusseerde toegang tot wetenschappelijke *Handbooks*. Wij nemen ons voor om een afzoekbare kaart te maken, die de belangrijke termen en hun relaties weergeeft alsmede een passende interface is naar een elektronisch in plaats van papieren *Handbooks*. Ook presenteren wij hier de algemene achtergrond van het zoeken naar en binnen boeken en over de structuren die voor indexerings- en classificatiedoelen nodig zijn.

In Hoofdstuk 3 benoemen wij onze eerste onderzoeksvraag. Dit is de vraag naar de vereisten waaraan een kaart moet voldoen die geschikt is voor menselijk bladeren en tevens een geraamte vormt voor gefocusseerde toegang tot een wetenschappelijk *Handbook*. Op basis van deze vereisten, passen wij het zo verkregen model toe op ons onderzoekscorpus: het domein van logica en taal. Wij presenteren dan onze kaart die wij de *Logic and Language Links*, of afgekort LoLaLi kaart noemen. Wij gaan in hoe zij tot stand is gekomen en wat haar onderdelen zijn, inclusief de hiërarchische en niet-hiërarchische relaties tussen topics. Als corpus voor ons onderzoek nemen wij het *Handbook of Logic & Language* onder redactie van Van Benthem en Ter Meulen.

In Hoofdstuk 4 behandelen we de tweede onderzoeksvraag: hoe presenteren we deze kaart aan lezers? Om te beginnen geven wij een overzicht van de bekende visualisatietechnieken voor semantische structuren zoals de onze. Daarna beschrijven wij de interface die wij ontwikkeld hebben om het eindgebruikers mogelijk te maken om binnen de LoLaLi kaart te navigeren en te zoeken. Wij sluiten af met een rapportage van gebruikersstudies die wij onder twee groepen studenten uitvoerden.

In Hoofdstuk 5 gaan we verder met onderzoeksvraag drie. Wat zijn de passende tekstdoelen in het *Handbook* om daadwerkelijk gefocusseerd en topic bepaald koppelingen te maken tussen de kaart en de tekst? Wij concentreren ons op de selectie van tekstexcerpten die door middel van twee soorten geautomatiseerde tekstsegmentatietechnieken geselecteerd worden. De ene manier is structuur, de andere semantisch georiënteerd. De resultaten vergelijken wij met een handmatige annotatie van het onderhavige *Handbook*.

Gegeven de verzameling tekstexcerpten die wij in het vorige hoofdstuk hebben verkregen gaan wij in Hoofdstuk 6 na in hoeverre deze tekstdelen passen bij de topics van de kaart. Dit is onze vierde onderzoeksvraag. Wij bespreken de evaluatie van dit proces en komen met voorstellen hoe zoiets aangepakt moet worden.

Uiteindelijk trekken wij in Hoofdstuk 7 onze conclusies en belichten richtingen voor verder onderzoek. Ons werk laat zien dat als wij dit *Handbook* voorzien met een domeinkaart, wij een alternatief pad om de tekst te benaderen hebben vormgegeven die door onze doelgroep gewaardeerd wordt.

Wij hebben geleerd van dit onderzoek dat een domeinkaart een informatief stuk gereedschap kan zijn en dat benoemde relaties zoals wij die gebruiken zowel sterke als zwakke kanten hebben. Aan de ene kant kunnen zij niet-expert gebruikers door de tekst en het vakgebied leiden, aan de andere kant hangt het begrip af van actuele kennis van de gebruiker wat kan leiden tot lastige interpretaties. De gebruikersstudies leerden ons

dat sommige kenmerken van de interface verbeterd moeten worden en dat sommige soorten gebruikers meer profijt hebben van de LoLaLi omgeving dan anderen.

Onze conclusie is dat een *topics*-kaart de basis kan vormen voor een elektronische omgeving waar een variëteit van informatiezoekgedrag ondersteund wordt en topic gestuurde toegang tot een tekst mogelijk wordt gemaakt. Naar onze mening zal echter het idee van een alles omvattende domeinkaart los moeten worden gelaten ten bate van beperktere kaarten die hetzij handmatig, hetzij (semi)automatisch geëxtraheerd worden uit de onderhavige tekst. Deze kaarten zouden specifiek toegang tot de tekst moeten bewerkstelligen en het onderwerpsdomein representeren volgens de interpretatie van de auteur(s). Deze kleiner kaarten kunnen dan op een dusdanige wijze aan elkaar verbonden worden dat gebruikers op een eenduidige manier kunnen navigeren, d.w.z. bladeren en zoeken, in een verzameling gerelateerde, maar onderling onafhankelijke bronnen.