Universiteit
Leiden
The Netherlands

Science maps for information retrieval

Bascur Cifuentes, J.P.

Citation
Bascur Cifuentes, J. P. (2026, January 21). *Science maps for information retrieval*. Retrieved from https://hdl.handle.net/1887/4287774

Version:	Publisher's Version
License:	Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden
Downloaded from:	https://hdl.handle.net/1887/4287774

Note: To cite this publication please use the final published version (if applicable).

Summary in Dutch

Science maps zijn een veelgebruikt hulpmiddel in scientometrische analyses. Een van hun belangrijkste voordelen is dat ze de structuur van gegevens zichtbaar maken, wat voor een analist zowel onbekende academische onderwerpen kan onthullen als eigen blinde vlekken kan blootleggen. Deze sterke punten maken ze goed geschikt voor Information Retrieval (IR), en onderzoekers maken hier dan ook vaak gebruik van. Toch richt het meeste bestaande onderzoek naar science maps zich op de nauwkeurigheid van onderwerpdetectie, terwijl veel minder aandacht is besteed aan hun mogelijkheden voor IR. Dit proefschrift behandelt deze leemte door één overkoepelende vraag te onderzoeken: Wat is de effectiviteit van science maps voor IR, en hoe kunnen we deze verbeteren?

Het proefschrift bestaat uit een inleiding, vier hoofdstukken (Hoofdstukken 2–5) die deelvragen van de overkoepelende vraag beantwoorden, en een conclusie. Elk van deze vier hoofdstukken is gebaseerd op een peer-reviewed publicatie.

Hoofdstuk 1: Inleiding

Dit hoofdstuk definieert wat science maps zijn en met welke IR-taken ze verband houden. Het plaatst het onderzoek binnen de bredere academische literatuur, introduceert de onderzoeksvragen die in latere hoofdstukken worden behandeld, en presenteert de aanvullende bijdragen van het proefschrift naast de onderzoeksvragen.

Hoofdstuk 2: Een interactieve visuele tool voor wetenschappelijke literatuurzoektocht: voorstel en algoritmische specificatie

Dit hoofdstuk behandelt de vraag "Hoe kunnen science maps worden ontworpen ter ondersteuning van IR?" door een tool voor te stellen die science maps integreert met een interactief retrievalproces. De tool stelt gebruikers in staat om iteratief een set documenten te verspreiden over clusters en vervolgens geselecteerde clusters te verzamelen tot een verfijnde subset (het 'scatter-gather' paradigma). Daarnaast hebben we een algoritme ontwikkeld om clusters te visualiseren door lege ruimte te minimaliseren en tegelijkertijd de betekenisvolle afstanden tussen de clusters te behouden.

Hoofdstuk 3: Academische IR met behulp van citatieclusters: diepgaande evaluatie op basis van systematische reviews

Dit hoofdstuk behandelt de vraag "Hoe effectief zijn science maps bij het uitvoeren van systematische reviews?". In de evaluatie hebben we IR gemodelleerd als een iteratief proces van het selecteren van clusters en het genereren van subclusters, met verschillende gebruikersmodellen die uiteenlopende voorkeuren hadden voor recall en precisie. De resultaten toonden aan dat science maps beter presteerden dan de booleaanse zoekopdrachten bij ongeveer de helft van de reviews. Dit wijst erop dat science maps het best gebruikt kunnen worden als aanvulling op, in plaats van als vervanging van, andere zoekhulpmiddelen.

Hoofdstuk 4: Welke onderwerpen worden het best weergegeven door science maps? Een analyse van clustereffectiviteit voor citatie- en tekstsimilariteitsnetwerken

Dit hoofdstuk behandelt de vraag "Vertegenwoordigen science maps sommige onderwerpen beter dan andere?" met behulp van Medical Subject Headings als onderliggend systeem voor onderwerpen. We hebben de effectiviteit van clusters per onderwerp afzonderlijk gemeten en vervolgens per themacategorie geaggregeerd. De best geclusterde categorieën waren Organismen en Ziekten, terwijl Geografische entiteiten het minst goed werden weergegeven. Een onderwerp werd als goed

geclusterd beschouwd als de bijbehorende documenten geconcentreerd waren in enkele clusters en die clusters weinig niet-gerelateerde documenten bevatten. We hebben de evaluatie uitgevoerd op science maps die gebouwd zijn zowel op citaties en inhoudelijke overeenkomst tussen documenten, op drie verschillende granulariteitsniveaus. Uit onze analyse bleek dat elke categorie vergelijkbare clustereffectiviteit vertoonde op beide typen science maps.

Hoofdstuk 5: Gebruik van diverse databronnen om te sturen welke onderwerpen verschijnen in een wetenschapskaart

Dit hoofdstuk behandelt de vraag "Hoe kan de weergave van specifieke onderwerpen in een science map worden verbeterd?" door de prestaties van onderwerpen te vergelijken tussen science maps die zijn opgebouwd met verschillende databronnen. We hebben acht soorten bronnen werden vergeleken: inhoudelijke overeenkomst, citaties, co-auteurschap, octrooien, beleidsdocumenten en verschillende vormen van sociale media. De evaluatiemethode lijkt op die van hoofdstuk 4, maar is aangepast om de vergelijking van een groter aantal science maps mogelijk te maken. Een belangrijke aanpassing was dat de prestaties van een onderwerp alleen werden vergeleken bij granulariteiten waarbij de verschillende kaarten hetzelfde aantal onderwerpclusters opleverden. De resultaten toonden aan dat verschillende bronnen konden verschuiven welke themacategorieën het best presteerden, maar dat de algehele clusterkwaliteit vaak afnam bij het gebruik van andere bronnen dan tekst- en citatienetwerken. Dit kwaliteitsverlies kon echter worden beperkt door databronnen te combineren.

Hoofdstuk 6: Conclusie

Dit hoofdstuk vat de bevindingen van de onderzoeksvragen samen en beantwoordt de overkoepelende onderzoeksvraag. Het benadrukt dat science maps effectief kunnen bijdragen aan IR, vooral in combinatie met andere tools, en dat hun geschiktheid verschilt per academisch onderwerp. Het hoofdstuk schetst ook richtingen voor toekomstig onderzoek, waaronder hoe betere prestaties kunnen worden bereikt met diverse databronnen, het gebruik van grote taalmodellen, en het belang van prototyping en software sustainability.