

Referenzen und Zitationen zur Unterstützung der Suche in SOWIPORT

Frank Sawitzki, Maria Zens, Philipp Mayr

GESIS – Leibniz-Institut für Sozialwissenschaften
D-50667 Köln

E-Mail: {frank.sawitzki, maria.zens, philipp.mayr}@gesis.org

Zusammenfassung

In SOWIPORT werden anhand von automatisch verlinkten Referenzen und Zitationen seit kurzem neue Suchmodi zur Unterstützung explorativer Suche angeboten. Literaturreferenzen werden dazu über spezifische Matching-Verfahren mit den Dokumenten aus dem gesamten Datenbestand verlinkt. Damit ist ein erster wichtiger Schritt in Richtung eines navigierbaren Zitationsindex für die deutschsprachigen Sozialwissenschaften getan.

Abstract

New search modes have been introduced recently in SOWIPORT to support the exploratory search on the basis of automatically linked references and citations. References are therefore linked to the complete data corpus via specific matching procedures. With this approach we have done a major step towards a navigable citation index in the field of the German social sciences.

1 Einführung

SOWIPORT (<http://www.sowiport.de>) ist ein Informationsportal für die Sozialwissenschaften mit steigenden Nutzerzahlen. Der enthaltene Bestand umfasst 19 Datenbanken mit ca. 7,2 Mio. Einträgen unterschiedlicher Informationstypen¹, dabei stellen Literaturnachweise das Gros der Datensätze. Aus den Datenbanken der nationallizenzierten CSA-Datenbanken sind ca. 8 Mio. Referenzen innerhalb der Sozialwissenschaften verfügbar.

¹ Literatur, Institutionen, Fachzeitschriften, Veranstaltungen, Sammlungen/Netzwerke, Forschungsprojekte, Studienbeschreibungen.

Diese Literaturreferenzen können künftig als zusätzliche Dokumentmetadaten in SOWIPORT angezeigt werden. Darüber hinaus können die im System enthaltenen Dokumente über die Referenzen automatisch verknüpft werden. Ziele der Verknüpfung sind a) die Unterstützung explorativer Suchen im Portal, b) die Verbesserung von Retrieval-Ergebnissen über zusätzliche Ranking-Optionen sowie c) der Aufbau eines navigierbaren Zitationsindex für die bibliometrische Forschung. Hervorzuheben ist, dass in letzterem auch die für die Sozialwissenschaften sehr wichtigen Monographien Berücksichtigung finden.

2 Zitationen für die Informationssuche

Referenzen haben eine zentrale Funktion für die Konstitution wissenschaftlicher Diskurse (Hyland 1999), sie ordnen das zitierende Werk in einen Forschungskontext ein und belegen die Rezeptions- und Wirkungsgeschichte des zitierten Werks. Die wissenschaftlichen Disziplinen verfahren dabei sehr unterschiedlich (Bornmann/Daniel 2008), die Stile und Verfahren der Referenzierung sind dabei Teil der Fachkultur. Howard D. White plädiert dafür, Zitationen als „a complex communicative process with syntactic, semantic, and pragmatic variables“ zu sehen (White 2004: 112).

In einem Portal wie SOWIPORT unterstützen Referenzen und die daraus aggregierten Zitationen das Information Retrieval in mehrfacher Hinsicht. Ausgangspunkt sind die von Marcia Bates so genannten Suchtaktiken des „footnote chasing“ und „citation searching“, also der Verkettung („chaining“) einer Publikation mit zitierten und zitierenden Arbeiten (Bates 1989). *Impact*: Mit der Berechnung von Zitationshäufigkeiten kann der Impact von Publikationen ermittelt werden, der wiederum hilft, wichtige Publikationen und Standardwerke zu ermitteln und bevorzugt anzuzeigen („forward chaining“ bei Bates). *Ähnliche Dokumente*: Über die Rekonstruktion disziplinärer Diskurse und Wirkungspfade können „ähnliche Dokumente“ z.B. über sog. Co-Citations ermittelt werden. Diese und weitere Suchtaktiken der explorativen Suche können Nutzer in ihrem Informationsverhalten unterstützen, sofern die notwendigen Informationen als „clickable paths“ vom System angeboten werden. Insbesondere der letzte Punkt ist wichtig für ein Angebot, das vor allem Wissenschaftlerinnen und Wissenschaftler in ihrer Forschungsarbeit unterstützen soll.

3 Referenzextraktion und -matching

SOWIPORT beinhaltet u.a. Literaturinformationen aus den Nationallizenzen für die englischsprachigen sozialwissenschaftlichen CSA-Datenbanken sowie dem Open Access Repository SSOAR. In den Lieferungen einiger CSA-Datenbanken² sind bereits strukturierte Informationen (Titel, Autoren, Publikationsjahr, Quelle) über in den Dokumenten enthaltene Referenzen vorhanden. Für die Datenbank SSOAR werden die Referenzen aus vorliegenden Volltexten mittels einer angepassten Version des Programms PARSCIT automatisch extrahiert (Haddou ou Moussa/Mayr 2012). Die Referenzen werden in einem ersten Verarbeitungsschritt in einem Referenzindex gespeichert und über die Angaben des Titels und des Erscheinungsjahres mit den Haupteinträgen aus dem Literaturkatalog verglichen, siehe Abb.1.

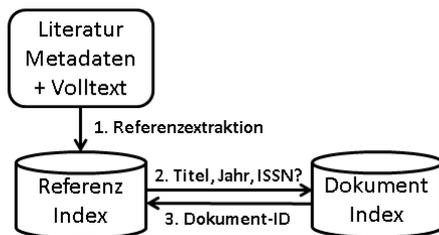


Abbildung 1:
Referenzdaten werden aus den Lieferungen extrahiert und die Entitäten Titel, Jahr, ISSN zwischen den Referenzen und den im Dokumentindex enthaltenen Dokumenten von SOWIPORT verglichen.

Ist dies nicht erfolgreich, erfolgt ein weiterer Abgleich über den Titel mit einer Unschärfefunktion (Fuzzy-Search), um den Recall zu erhöhen. Gleichzeitig werden Erscheinungsjahr und ISSN hinzugenommen, um die Precision nicht absinken zu lassen. Bei einer Übereinstimmung wird durch Speichern der Dokument-ID in den Referenzindex eine Verknüpfung zwischen Referenzeintrag und referenziertem Dokument hergestellt.

Bei einer periodischen Aktualisierung aller Datenbanken werden in einem weiteren Verarbeitungsschritt die Referenzinformationen inklusive der Verlinkung den Dokumenten hinzugefügt, bzw. Zitationen mittels der vorliegenden Verknüpfungen aus dem Referenzindex ermittelt, s. Abb. 2.

² Worldwide Political Science Abstracts (WPSA), Sociological Abstracts (SA) und Social Services Abstracts (SSA)

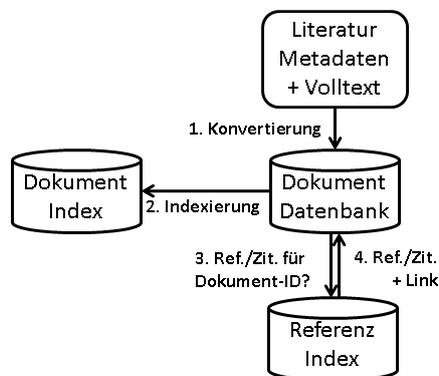


Abbildung 2:

Nach einer Konvertierung der Metadaten werden die Informationen in eine Dokument-Datenbank gespeichert und ein Index aufgebaut. Anschließend wird der Referenzindex abgefragt, ob Referenzen oder Zitationen für ein Dokument vorliegen. Diese Informationen werden den Dokumenten inklusive einer evtl. gefundenen Verlinkung in der Datenbank angefügt.

Zudem werden die Referenz- und Zitationsangaben in die Darstellung übernommen und, falls verfügbar, als Verknüpfungen angeboten, s. Abb. 3.

Role of Socialization in Explaining Social Inequalities in Health ▲
 Autor: Marmot, Michael; Singh-Manoux, Archana
 Erscheinungsjahr: 2005; Informationstyp: Literatur; Dokumenttyp: Zeitschriftenaufsatz
 Datenbank: Sociological Abstracts (CSA)
 Referenzen: 51; Zitiert von (Beta-Version): 6

Quelle	Serialtitel: Social Science & Medicine, vol. 60, no. 9, pp. 2129-2133, May 2005, ISSN: 0277-9536
Inhalt	This paper argues that social selection, materialist/structural &
Referenzen	Adler, N E; Boyce, T; Chesney, M A (1994): Socioeconomic status and health. The challenge of the gradient 15-24. Berkman, L F; Glass, T; Brisette, I (2000): From social integration of health: durkheim in the new millennium Social Science & Medicine 51, 843-857. Bosma, H; Van de Mheen, H; Mackenbach, J P (1999): Social class in childhood and general health in adulthood: questionnaire study of contribution of psychological attributes British Medical Journal 318, 18-22.
Zitiert von	Elstad, Jon Ivar (2010): Indirect health-related selection or social causation? Interpreting the educational differences in adolescent health behaviours Ward, Michael M; Guthrie, Lori C; Butler, Stephen C (2009): Time perspective and socioeconomic status: A link to socioeconomic disparities in health?

Abbildung 3:

Darstellung von Referenzinformationen unter dem Eintrag „Referenzen“ innerhalb der Dokumentdarstellung des Informationsportals SOWI-PORT. Hervorgehoben gekennzeichnet sind Einträge, zu denen Verknüpfungen vorliegen. Unter „Zitiert von“ werden die ermittelten Zitationen des Dokuments aufgelistet.

Mittels des beschriebenen Matching-Verfahrens können in unserer ersten Implementierung ca. 30% der enthaltenen Referenzen den zugehörigen Dokumenten zugeordnet werden. Über das Verfahren konnten somit über 2,5 Mio. zusätzliche Referenz- bzw. Zitationsverknüpfungen für die explorative Suche bereitgestellt werden. Die Überprüfung der Qualität der Verknüpfungen mit einer Zufallsstichprobe von ca. 400 Referenzen ergab eine Rate von korrekten Erkennungen von 95%. Ein Vergleich dieses Ergebnisses mit denen anderer Indizes wie z.B. des Social Science Citation Index wäre wünschenswert, ist aber derzeit nicht möglich, da dieses sein Verfahren und die damit erzielten Erkennungsraten nicht offenlegt.

4 Ausblick

Weitere Schritte bei der Entwicklung von SOWIPORT sind die sukzessive Verbesserung von Recall und Precision der Matching-Verfahren, Dublettenkontrolle und Autor-Disambiguierung sind dabei zentral. Hierzu werden wir die Einbeziehung weiterer Informationen aus den Referenzen prüfen (insb. die Autor-Affiliation). Übergreifendes Ziel im Projekt ist es, auf Basis der Referenzen und Zitationen weitere Mehrwertdienste zu entwickeln, die Wissenschaftler und Wissenschaftlerinnen bei der Suche, der Bewertung und dem Explorieren von Treffermengen individuell unterstützen. Neben der Verknüpfung von Publikationen über Referenzen und Zitationen sollen in SOWIPORT weitere Verlinkungen realisiert werden. Dabei bietet es sich an, Forschungsprojekte (Datenbank SOFIS) mit den zugehörigen Publikationen in den Literaturdatenbanken oder Studiendaten (z.B. aus dem Datenbestandskatalog) mit Publikationen zu verknüpfen. Beides wird derzeit erprobt, prototypische Ergebnisse stehen bereits zur Verfügung (Boland 2012).

5 Literaturverzeichnis

Bates, M. (1989). The Design of Browsing and Berrypicking Techniques for the Online Search Interface. *Online Review* 13, 5, pp. 407-424.

Boland, K.; Ritze, D.; Eckert, K.; Mathiak, B. (2012). Identifying References to Datasets in Publications. *LNCS* 7489, pp. 150-161.

Bornmann, L; Daniel, H. (2008). What do citation counts measure? A review of studies on citing behavior. *Journal of Documentation* 64, 1, pp. 45-80.

Haddou ou Moussa, K.; Mayr, Ph. (2012). Automatische Referenzextraktion mit PARSCIT. *Social Media and Web Science: das Web als Lebensraum; Proceedings; 2. DGI-Konferenz 2012*, pp. 425-428.

Hyland, K. (1999). Academic attribution: citation and the construction of disciplinary knowledge. *Applied Linguistics* 20, 3, pp. 341-367.

White, H. D. (2004). Citation Analysis and Discourse Analysis Revisited. *Applied Linguistics* 25, 1, pp. 89-116.