

참조연결을 위한 인용정보 자동추출에 관한 연구

A Study on Automatic Extraction of Citation Information for Reference Linking

김 지 훈(Ji-Hoon Kim)

초 록

최근 인터넷의 빠른 성장과 전자출판으로 인해 정보에 접근하는 방법이 상당히 변화하고 있다. 이와 함께, 디지털 문서에서 적합한 정보를 결정하고 효율적으로 추출할 필요가 있을 것이다. 이 연구는 디지털 문서에서 여러 가지 정보를 추출하기 위해 템플릿 마이닝이 사용될 수 있음을 제시하고, 참조연결시스템과 관련된 연구를 고찰하였다. 또한 실제적으로 논문 샘플을 분석한 결과를 이용하여 참조정보를 위한 템플릿을 구축하고, 이 템플릿을 이용하여 수작업으로 논문 샘플을 테스트한 결과, 템플릿 마이닝의 이용은 인용 데이터베이스를 자동으로 만들 수 있는 가능성을 보여주었다.

ABSTRACTS

Recently, the way information is accessed has changed significantly because of the rapid growth of the Internet and a move toward electronic publishing. With that condition, there will be a need to efficiently determine and extract relevant information from digital documents. This study has proposed that template mining can be used for extracting different kinds of information from digital documents, and described research of reference linking system and then built template for citation information using analysis result of sample articles. Also, this study has shown the potential of template mining to automatically create citation databases.

키워드: 참조연결, 인용연결, 인용색인, 인용정보추출, 템플릿마이닝, 인용데이터베이스

이 논문은 2002년도 한국문헌정보학회 추계학술발표회(2002. 10. 11-12)에서 발표된 내용을 보완한 것임.

계명문화대학 문헌정보과 부교수(jhkim@km-c.ac.kr)

논문접수일자 2003년 2월 28일

게재확정일자 2003년 3월 10일

참고문헌

Andersen, P. M. and Huettner, A. K. 1994. Knowledge Engineering for the JASPER Fact Extraction System. *Integrated Computer-Aided Engineering*. 1(6):473-493.

Atkins, Helen. 1999. The ISI Web of Science - Links and Electronic Journals. *D-Lib Magazine*. 5(9).

<<http://www.dlib.org/dlib/september99/atkins/09atkins.html>>.

Atkins, Helen; Lyons, Catherine; Ratner, Howard; Risher, Carol; Shillum, Chris; Sidman, David and Stevens, Andrew. 2000. Reference Linking with DOIs: A Case Study. *D-Lib Magazine*. 6(2).

<<http://www.dlib.org/dlib/february00/02risher.html>>

Bergmark, Donna. 2000. Automatic Extraction of Reference Linking Information from Online Documents. Technical Report CSTR 2000-1821, Cornell University, Digital Library Research Group.

<<http://www.cs.cornell.edu/cdlrg/Reference Linking/extraction.pdf>>

Bergmark, Donna and Lagoze, Carl. 2001. Reference Linking the Web's Scholarly Papers. CSTR 2001-1835, Cornell University, Digital Library Research Group.

<<http://www.cs.cornell.edu/cdlrg/Reference Linking/LinkingTheWeb.pdf>>

Bollacker, Kurt D.; Lawrence, Steve and Giles, C. Lee. 1998. CiteSeer: An Autonomous Web Agent for Automatic Retrieval and Identification of Interesting Publications. In Katia P. Sycara and Michael Wooldridge (eds.). *Proceedings of the Second International Conference on Autonomous Agents*. New York: ACM Press. 116-123.

Cameron, R. D. 1997. A Universal Citation Database As a Catalyst for Reform in Scholarly Communication. *First Monday*. 2(4).

<http://www.firstmonday.dk/issues/issue2_4/cameron/index.html>

Caplan, Priscilla and Arms, William Y. 1999. Reference Linking for Journal Articles. *D-Lib Magazine*. 5(7/8).
<<http://www.dlib.org/dlib/july99/caplan/07caplan.html>>

Chowdhury, Gobinda G. 1999. Template Mining for Information Extraction from Digital Documents. *Library Trends*. 48(1): 182-208.

Constantino, M.; Morgan, R. G. and Collingham, R. J. 1996. Financial Information Extraction using Pre-defined User-definable Templates in the LOLITA System. *Journal of Computing & Information Technology*. 4(4):241-255.

Cronin, Blaise; Snyder, Herbert and Atkins, Helen. 1997. Comparative Citation Rankings of Authors in Monographic and Journal Literature: a Study of Sociology. *Journal of Documentation*. 53(3):263-273.

CrossRef: The Central Source for Reference Linking.
<<http://www.crossref.org>>

Ding, Ying; Chowdhury, Gobinda and Foo, Schubert. 1999. Template mining for the extraction of citation from digital documents. *Proceedings of Second Asian Digital Libraries Conference, National Taiwan University, November 8-9, 1999*.
<<http://www.cs.vu.nl/~ying/download/ad19911.pdf>>

Hitchcock, S.; Carr, L.; Harris, S.; Hey, J. M. N. and Hall, W. 1997. Citation Linking: Improving Access to Online Journals. In Robert B. Allen and Edie Rasmussen. (eds.). *Proceedings of the 2nd ACM International Conference on Digital Libraries*. New York: ACM. 115-122.
<<http://journals.ecs.soton.ac.uk/acmdl97.htm>>

Hitchcock, Steve. Carr, Les. Hall, Wendy. Harris, Steve. Proberts, Steve. Evans, David and Brailsford, David. 1998. Linking Electronic Journals: Lessons from the Open Journal Project. *D-Lib Magazine*. December 1998.
<<http://www.dlib.org/dlib/december98/12hitchcock.html>>

Hitchcock, Steve; Carr, Les; Jiao, Zhuan; Bergmark, Donna; Hall, Wendy; Lagoze, Carl and Harnad, Stevan. 2000. Developing Services for Open Eprint Archives: Globalisation, Integration and the Impact of Links. In *ACM Proceedings of Digital Libraries, San Antonio, Texas, June 2 - June 7, 2000*.
<<http://opcit.eprints.org/dl00/htdl00.pdf>>

Jacobs, P. and Ratt, L. F. 1990. SCISOR: Extracting Information from Online News. *Communications of the ACM*. 33(11):88-97.

Giles, C. Lee; Bollacker, Kurt D. and Lawrence, Steve. 1998. CiteSeer: An Automatic Citation Indexing System. In Ian Witten, Rob Akscyn, and Frank M. Shipman III, (eds). *Digital Libraries 98 - The Third ACM Conference on Digital Libraries, Pittsburgh, PA, June 23-26 1998*. Pittsburgh: ACM Press. 89-98.

Lawrence, Steve; Bollacker, Kurt and Giles, C. Lee. 1999. Indexing and Retrieval of Scientific Literature. *Eighth International Conference on Information and Knowledge Management, CIKM 99, Kansas City, Missouri, November 2-6*. 139-146.

Lawrence, Steve; Giles, C. Lee and Bollacker, Kurt. 1999. Digital Libraries and autonomous citation indexing. *IEEE Computer*. 32(6): 67-71.

Lawrence, Steve. 2001. Online or invisible?. *Nature*. 411(6837):521.

Lawson, M.; Kemp, N.; Lynch, M. F. and Chowdhury, G.G. 1996. Automatic extraction of citations from the text of English language patents: An example of template mining. *Journal of Information Science*, 22(6): 423-436.

NASA Astrophysics Data System.
<<http://adswww.harvard.edu>>

The NLM PubMed Project.
<www.ncbi.nlm.nih.gov/PubMed/overview.html>

An Open Journal Framework: Integrating Electronic Journals with Networked Information Resources. 1995.
<<http://journals.ecs.soton.ac.uk/flyer.html>>

Paice, C. D. 1981. The Automatic Generation of Literature Abstracts: an Approach Based on the Identification of Self-indicating Phrases. In R. N. Oddy, S. E. Robertson, C. J. Van Rijsbergen and P. W. Williams (eds.) Information Retrieval Research. London: Butterworths. 172-191.

Paskin, Norman. 1999. DOI: Current Status and Outlook. D-Lib Magazine. 5(5).

<<http://www.dlib.org/dlib/may99/5paskin.html>>

ResearchIndex.

<<http://www.neci.nec.com/~lawrence/researchindex.html>>

What is Reference Linking?.

<http://arabica.ecs.soton.ac.uk/ref_linking.html>

K C I