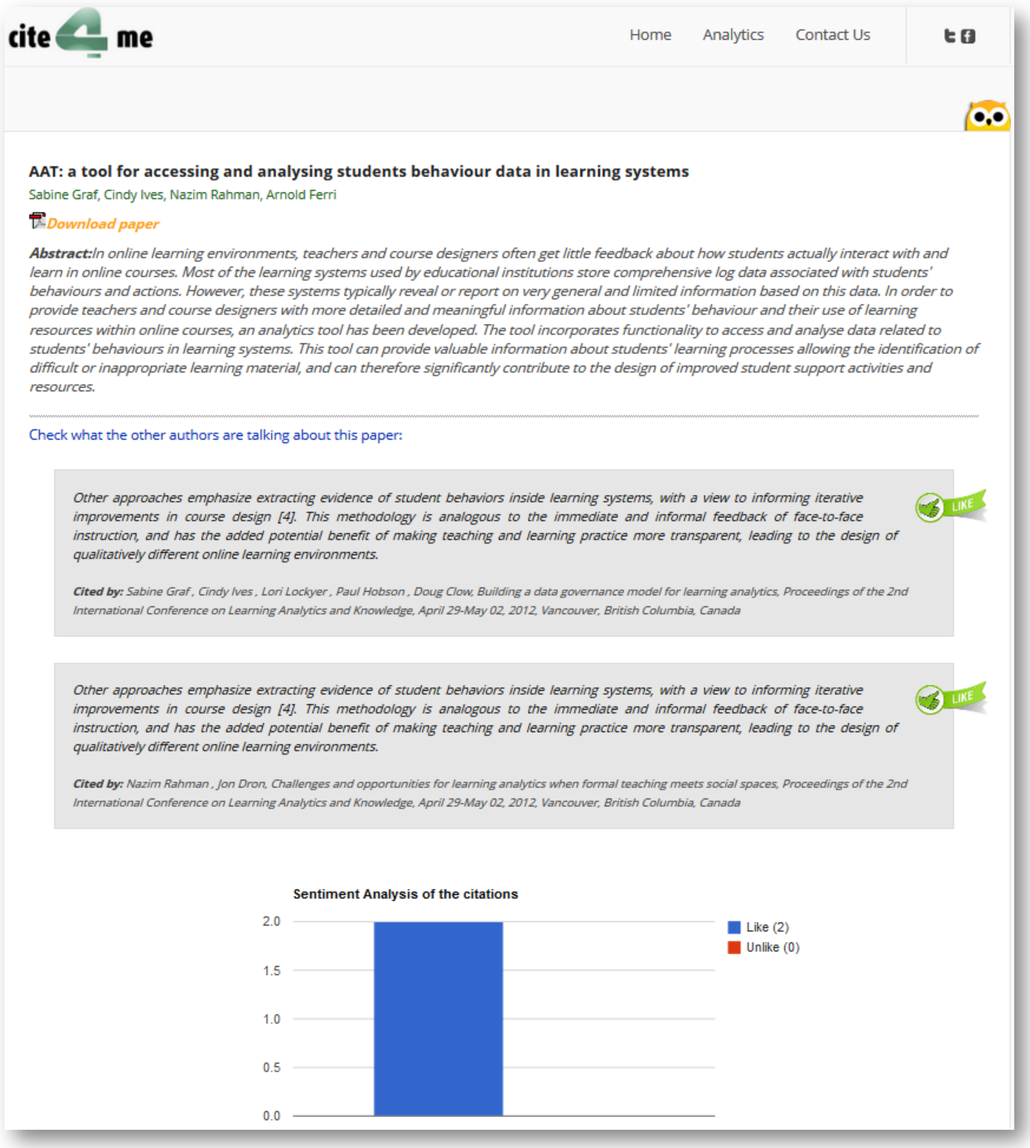
Cite4Me: A Semantic Search and Retrieval Web Application for Scientific Publications

Bernardo Pereira Nunes[1,2] , Besnik Fetahu[1], Stefan Dietze[1], Marco A. Casanova[2]

[1]Leibniz University Hannover (Germany), [2]Pontifical Catholic University of Rio de Janeiro (Brazil)



Cite4Me in action – http://www.cite4me.com

Abstract

Cite4Me is a Web application that leverages Semantic Web technologies to provide a new perspective on search and retrieval of bibliographical data.

The Web application presented in this work focuses on:

① Semantic recommendation of papers

② Novel semantic search & retrieval of papers

③ Data interlinking of bibliographical data with related data sources from LOD

④ Innovative user interface design

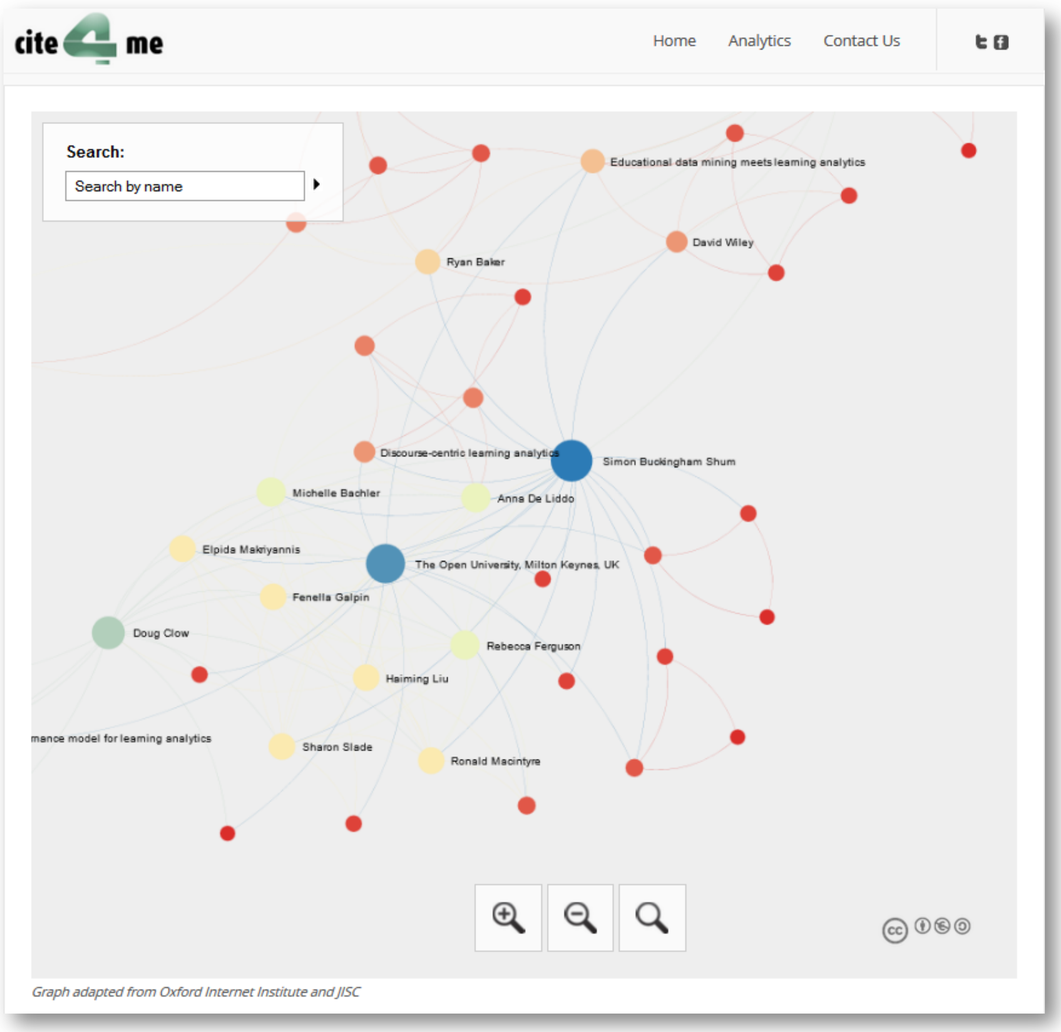
⑤ Sentiment analysis of extracted paper citations

Cite4Me - The Application

Cite4Me implements semantic and co-occurrence-based methods to search and retrieve academic papers and suggest related work in a user-friendly interface that assists users in exploring relationships between authors, institutions, papers and query terms.

Exploratory Search

The exploratory search or graph search component assists users to discover related work, people and institutions that are working on a specific topic.



Graph adapted from Oxford Internet Institute and JISC

Semantic Search

After running an annotation process, the relatedness score between the enriched concepts (DBpedia entities) found in the user query terms and the publications' content are computed and ranked. The relatedness score is computed based on the *tf-idf* score for the entities found in the publications' content. The ranking of the retrieved documents is based on the sum of the *tf-idf* scores of the matching concepts.

Paper recommendation

Another important feature of Cite4Me and which differentiates it from similar tools is the semantic paper recommendation. Given a scientific publication, the tool recommends a related paper based on a score calculated according to direct and lateral relationships between the publication of interest and the remaining papers in our corpus.

Evaluation

Currently, Cite4Me is linked to a dataset (LAK Dataset) which contains semi-structured research publications from the ACM Digital Library (under a special license) and other public datasets .The dataset contains 315 full papers along with their descriptive metadata while new publications are added continuously. Metadata as well as the full text body are freely available in a variety of formats, including RDF accessible via a public SPARQL endpoint.

Acknowledgements: CNPq and FAPERJ