Macademia: Semantic Visualization of Research Interests

Shilad Sen and the Macademia Development Team[ψ]
Department of Mathematics, Statistics, and Computer Science
Macalester College
ssen@macalester.edu

ABSTRACT
The Macademia website promotes faculty collaboration and research by visualizing research interests as a dynamic "constellation." Semantic similarity inference algorithms power the site's visualization, allowing users to spatially browse people connected through similar research interests.

Author Keywords
User interfaces, tagging, semantic similarity, visualization.

ACM Classification Keywords
H5.2. Information interfaces and presentation (e.g., HCI)

General Terms
Algorithms, Design

INTRODUCTION
Researchers who collaborate with other people produce more research [3]. Recent social technical tools make it easier to identify potential collaborators. For example, users can search for a researcher's profile by name on websites such as academia.edu and ResearchGate. However, identifying researchers by research interest may be difficult because researchers use different vocabularies for similar concepts [5,6]. Should somebody search for "critical theory," "literary critisicm," or "hermeneutics?"

The Macademia website (http://macademia.macalester.edu) supports searching and browsing for people by research interest by incorporating two novel design elements. First, the website infers and displays semantic relationships between research interests. For example, Macademia knows that "online communities" is more closely related to "web2.0" than "social psychology." Second, Macademia visualizes people and research interests as a dynamic constellation that captures semantic relationships.

MACADEMIA INTERFACE
The Macademia visualization centers on a person (Figure 2) or research interest (Figure 3). In both cases, two rings of *nodes* (people or interests) encircle the center node. Edges join semantically related nodes. When a user clicks a node, the constellation spins to re-center around that node.

Person-centered visualization: The person's research interests appear in the first ring, and people with semantically related interests appear in the second ring. Hovering over a person describes the interest connections between the center node and the second-ring person.

IUI'11, February 13–16, 2011, Palo Alto, California, USA.
ACM 978-1-4503-0419-1/11/02.

Interest-centered visualization: Semantically related interests appear in the first ring, and people with those interests appear in the second ring.

Macademia offers other features to support users:

- **Tooltips** display detailed information about a person or interest when a user hovers over the node and highlight semantic connections to that node.
- **School filtering** limits the search results and visualization to faculty at particular institutions.
- **Dynamic profiles** update the visualization and semantic relationships as soon as users create or change their profile.
- **Collaborator requests** allow users to solicit specific types of expertise for a research project.
- **A density slider** shows fewer or more nodes on the graph to accommodate different display resolutions.

MACADEMIA SEMANTIC SIMILARITY ALGORITHM
Macademia visualizes semantic relationships between research interests. We considered the co-occurrence based semantic inference algorithms used in tagging systems [1]. However, our pilot studies suggested that research interests overlapped much less frequently than traditional tags. Thus, we adopted an approach inspired by [4] (Figure 1).

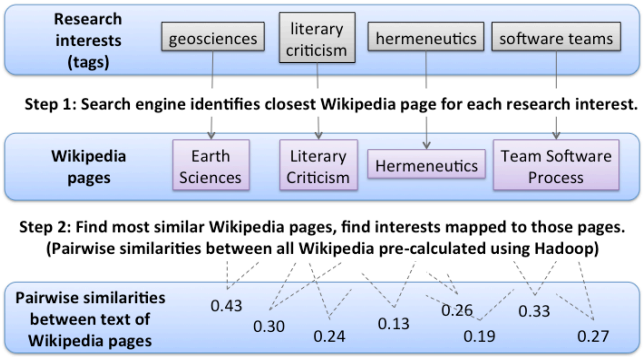

Figure 1. Macademia calculates the semantic similarity between interests by mapping them to Wikipedia pages.

ψ Henry Charlton, Ryan Kerwin, Jeremy Lim, Brandon Maus, Nathaniel Miller, Alex Schneeman, Anthony Tran, Megan R. Naminski (Macalester College) {hcharlton|rkerwin|jlim|bmaus|nmiller4|aschneem|atran}@macalester.edu, mnaminski@gmail.com; Ernesto Nunes (University of Minn.) enunes@cs.umn.edu; E. Isaac Sparling (Socialcast) isaac.sparling@gmail.com.

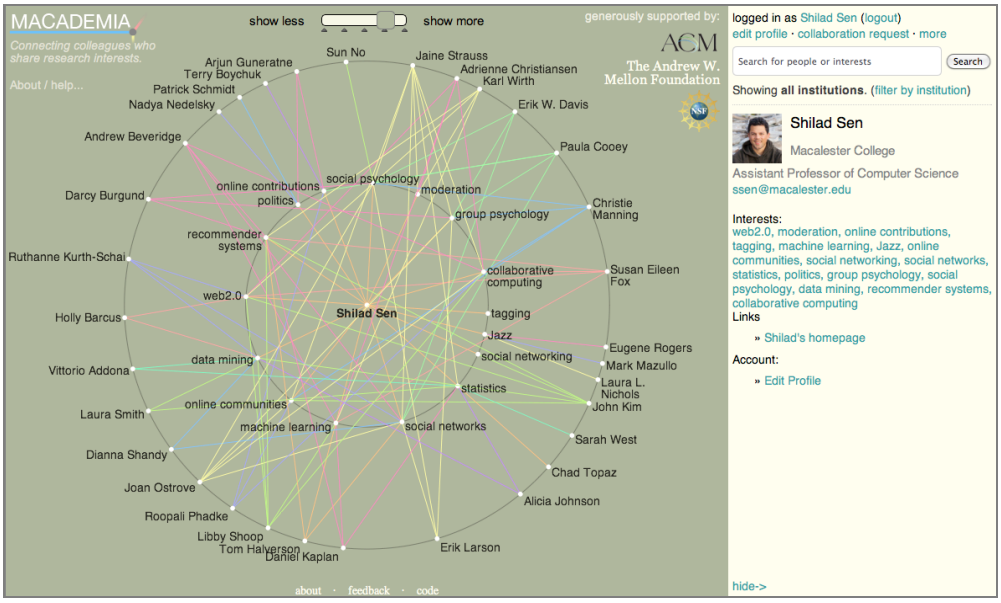

Figure 2. The main Macademia screen. The constellation centers around a researcher. Their interests appear in the first ring, and people with semantically related interests appear in the second ring. Clicking on a person or interest recenters the graph. A tooltip describes the connection between the center node and another person hovering over the second person.

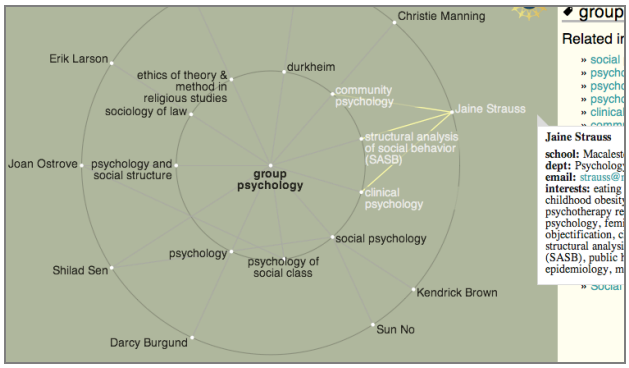

Figure 3. The interest-centered Macademia visualization. Related interests appear in the first ring, and researchers who have those interest lie in the second ring

First, Macademia finds the Wikipedia page p_1 closest to a described research interest i_1[1]. Second, it finds pages p_2 that are most semantically similar to p_1. Macademia measures semantic similarity using Tf-Idf document similarity between Wikipedia pages. Third, it reverse maps each p_2 to research interests i_2, if they exist in the system.

Mapping research interests to Wikipedia pages confers two benefits. First, Wikipedia pages offer rich textual descriptions that support semantic similarity measures better than simple tags. Second, Macademia pre-computes and stores pair wise similarities between all Wikipedia pages using the Hadoop parallel processing framework as described in [2]. Pre-computing these relationships enables Macademia to adapt to new research interests in constant time, regardless of the number of interests in the system.

MACADEMIA USAGE

Macademia currently supports over 300 faculty at the Associated Colleges of the Midwest. It will soon be open to all researchers. We hope to use the site as a platform to study semantic tagging systems in the future.

THE ACKNOWLEDGEMENTS

Macademia has been generously funded through grants from the Associated Colleges of the Midwest, the Andrew W. Mellon Foundation, NSF grant IIS-0964697, and Macalester College.

REFERENCES

[1] C. Cattuto, D. Benz, A. Hotho, and G. Stumme. Semantic grounding of tag relatedness in social bookmarking systems. *The Semantic Web-ISWC 2008*, :615-631, 2010.

[2] T. Elsayed, J. Lin, and D.W. Oard. Pairwise document similarity in large collections with mapreduce. In *Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics on Human Language Technologies: Short Papers*, pages 265-268. Association for Computational Linguistics, 2008.

[3] S. Lee and B. Bozeman. The impact of research collaboration on scientific productivity. *Social Studies of Science*, 35(5):673, 2005.

[4] A. Marchetti, M. Tesconi, F. Ronzano, M. Rosella, and S. Minutoli. Semkey: a semantic collaborative tagging system. In *Workshop on Tagging for Social Information Organization at WWW*, pages 8-12. Citeseer, 2007.

[5] K. McKenzie and N.S. Crowcroft. Describing race, ethnicity, and culture in medical research. *Bmj*, 312(7038):1054, 1996.

[6] C. Shirky. Shirky: ontology is overrated--categories, links, and tags. *Clay Shirky's Internet Writings*, 242009.

[1] We use a Google site-specific search for the research interest.