



ABCDE



Scientific Report

First name / Family name

Jodi Schneider

Nationality

USA

Name of the *Host Organisation*

INRIA

First Name / family name
of the *Scientific Coordinator*

Fabien Gandon

Period of the fellowship

08/12/2013 to 08/06/2015



I – SCIENTIFIC ACTIVITY DURING YOUR FELLOWSHIP

Overall from my ABCDE fellowship I have 6 papers – 1 published conference paper, 3 workshop papers, 1 experience report, and 1 invited paper. I have also written 4 unpublished manuscripts, 3 unpublished abstracts, and a grant proposal. I have given 10 talks at a variety of events – 1 tutorial, 1 panel presentation, 2 lightning talks, 3 long talks for conferences without proceedings, and 3 invited talks.

During my fellowship I worked on the following three main projects.

Argument mining: New contacts at INRIA, including Serena Villata and Elena Cabrio, have been instrumental in helping me define my research agenda on argumentation mining. During the fellowship I have drafted a review of 150+ papers in the area with the support of mentor Fabien Gandon (see manuscripts, below). In summer 2014, for the first time, argumentation mining was featured at a scientific event; in fact, 5 different events related to argumentation mining took place between May and July 2014, and I was fortunate to attend all of them (see publications: First Workshop on Argumentation Mining, Frontiers and Connections between Argumentation Theory and Natural Language Processing; talks: ArgDiaP, Arguing on the Web 2.0, SICSA). Besides my own papers, a manuscript started at the SICSA workshop (see manuscripts below) will also be released this year.

Evidence curation: Drawing on my work in argumentation mining, I am studying how research findings are evaluated and translated into practice. This work first took shape in a talk I gave on “evidence informatics” (SICSA 2014) and a submission to the European Conference on Argumentation. In collaboration with a National Library of Medicine project on medication safety, I am studying how groups of health professionals collaboratively curate and make sense of evidence on drug-drug interactions. So far we have published 3 workshop papers, and have a 4th paper in progress.

Newcomers and information quality in Wikipedia: Wikipedia’s sustainability depends on volunteers who can uphold the encyclopedia’s high quality standards: this is increasingly pressing as Wikipedia finds that volunteers are leaving faster than they can be replaced. I have one manuscript related to newcomers and information quality in progress: I am currently revising “Socialization of newcomers through quality-control discussions in an open collaboration system”, which extends findings from my dissertation, for publication.

Further, I published a conference paper investigating Wikipedia’s process for evaluating newcomers’ draft articles (OpenSym 2014). This work, with two new collaborators, also resulted in a small grant for Bluma Gelley (NYU) on using machine learning to determine whether content meets Wikipedia’s notability standard. Further, a visit to INRIA by new collaborator Aaron Halfaker (Research and Data team, Wikimedia Foundation Analytics group) laid the groundwork for wider collaboration around French Wikipedia and French DBpedia.

The fellowship has been instrumental in enabling me to collaborate on new projects, distinct from my dissertation research, and to network with researchers throughout Europe and North America.



II – PUBLICATION(S) DURING YOUR FELLOWSHIP

Full conference papers:

Jodi Schneider, Aaron Halfaker, Bluma Gelley. “Accept, decline, postpone: How newcomer productivity is reduced in English Wikipedia by pre-publication review”. *Proceedings of the 10th International Symposium on Open Collaboration (OpenSym, Wikipedia track)*, Berlin, Germany, August 2014.

ABSTRACT: Wikipedia needs to attract and retain newcomers while also increasing the quality of its content. Yet new Wikipedia users are disproportionately affected by the quality assurance mechanisms designed to thwart spammers and promoters. English Wikipedia’s Articles for Creation provides a protected space for drafting new articles, which are reviewed against minimum quality guidelines before they are published. In this study we explore how this drafting process has affected the productivity of newcomers in Wikipedia. Using a mixed qualitative and quantitative approach, we show how the process’s pre-publication review, which is intended to improve the success of newcomers, in fact decreases newcomer productivity in English Wikipedia and offer recommendations for system designers.

Workshop papers associated with international conferences:

Jodi Schneider, Paolo Ciccarese, Tim Clark and Richard D. Boyce. “Using the Micropublications ontology and the Open Annotation Data Model to represent evidence within a drug-drug interaction knowledge base”. *4th Workshop on Linked Science 2014—Making Sense Out of Data (LISC2014)* at the 13th International Semantic Web Conference (ISWC 2014) Riva de Garda, Italy, October 19, 2014.

ABSTRACT: Semantic web technologies can support the rapid and transparent validation of scientific claims by interconnecting the assumptions and evidence used to support or challenge assertions. One important application domain is medication safety, where more efficient acquisition, representation, and synthesis of evidence about potential drug-drug interactions is needed. Potential drug-drug interactions (PDDIs), defined as two or more drugs for which an interaction is known to be possible, are a significant source of preventable drug-related harm. The combination of poor quality evidence on PDDIs, and a general lack of PDDI knowledge by prescribers, results in many thousands of preventable medication errors each year. While many sources of PDDI evidence exist to help improve prescriber knowledge, they are not concordant in their coverage, accuracy, and agreement. The goal of this project is to research and develop core components of a new model that supports more efficient acquisition, representation, and synthesis of evidence about potential drug-drug interactions. Two Semantic Web models—the Micropublications Ontology and the Open Annotation Data Model—have great potential to provide linkages from PDDI assertions to their supporting evidence: statements in source documents that mention data, materials, and methods. In this paper, we describe the context and goals of our work, propose competency questions for a dynamic PDDI evidence base, outline our new knowledge representation model for PDDIs, and discuss the challenges and potential of our approach.

Mathias Brochhausen, **Jodi Schneider**, Daniel Malone, Philip E. Empey, William R. Hogan and Richard D. Boyce “Towards a foundational representation of potential drug-drug interaction knowledge”. First International Workshop on Drug



Interaction Knowledge Representation (DIKR-2014) at the 5th International Conference on Biomedical Ontologies (ICBO 2014) Houston, Texas, USA, October 6, 2014.

ABSTRACT: Inadequate representation of evidence and knowledge about potential drug-drug interactions is a major factor underlying disagreements among sources of drug information that are used by clinicians. In this paper we describe the initial steps toward developing a foundational domain representation that allows tracing the evidence underlying potential drug-drug interaction knowledge. The new representation includes biological and biomedical entities represented in existing ontologies and terminologies to foster integration of data from relevant fields such as physiology, anatomy, and laboratory sciences.

Jodi Schneider, “Automated argumentation mining to the rescue: Envisioning argumentation and decision-making support for debates in open online collaboration communities”. *First Workshop on Argumentation Mining* at the 52nd Annual Meeting of the Association for Computational Linguistics (ACL 2014), Baltimore, Maryland, June 26, 2014.

ABSTRACT: Argumentation mining, a relatively new area of discourse analysis, involves automatically identifying and structuring arguments. Following a basic introduction to argumentation, we describe a new possible domain for argumentation mining: debates in open online collaboration communities. Based on our experience with manual annotation of arguments in debates, we envision argumentation mining as the basis for three kinds of support tools, for authoring more persuasive arguments, finding weaknesses in others’ arguments, and summarizing a debate’s overall conclusions.

Conference Experience report:

Krystian Samp, Cedric Beuzit and **Jodi Schneider**. “Unifying the Shift and Narrow Strategies in Focus+Context Exploratory Search”. *Special Interest Group on Design of Communication (SigDOC)*. Colorado Springs, CO. September 27-28, 2014.

ABSTRACT: In this paper we discuss two existing exploration strategies – *Shift* and *Narrow* – employed by Focus+Context techniques, and how they are supported in the user interface of Saffron, a web-based system enabling exploration of academic topics, authors, and publications. The Shift strategy enables the user to shift focus between different resources while the Narrow strategy enables the user to narrow the focus. Current systems typically support only one of these approaches or include them as separate interaction modes. Saffron supports both strategies in a unified user interface. An initial user study indicates that participants use and appreciate both strategies being supported simultaneously.

International events with proceedings (not workshops or conferences):

Jodi Schneider. “An Informatics Perspective on Argumentation Mining”. In Proceedings of Frontiers and Connections between Argumentation Theory and Natural Language Processing. Bertinoro (Forlì-Cesena), Italy, July 21 – 25, 2014.

ABSTRACT: It is time to develop a community research agenda in argumentation mining. I suggest some questions to drive a joint community research agenda and then explain how my research in argumentation, on support tools and knowledge representations, advances argumentation mining.



Manuscripts pending:

Jodi Schneider, Serena Villata, Elena Cabrio, “Why did they post that argument? Communicative intentions of Web 2.0 arguments”. *Arguing on the Web 2.0* at the 8th ISSA Conference on Argumentation, Amsterdam, The Netherlands, June 30–July 1, 2014.

(Pending: manuscript invited for submission to post-conference proceedings)

ABSTRACT: The Web is an open platform where users are free to publish their own opinions, to discuss the latest news, to write reviews about a service or product. Given the growing number of online platforms allowing such interactions, it is becoming more and more important to support users in understanding the meaning of such conversations by identifying the context of the discourse, and the evidence and background knowledge required to evaluate the proposed arguments. In this paper, we suggest recording the rough context with communicative intentions to help classify arguments on the Web. We describe five communicative intentions: recreation, information, instruction, discussion, and recommendation. We suggest that this classification can help identify and analyze messages for the Argument Web. In particular, we discuss the suitable combinations of natural language processing techniques and argumentation-based reasoning to support users in understanding the discussions.

Jodi Schneider and Fabien Gandon. “A review of argumentation mining: how to structure and retrieve arguments from text”.

(Pending: manuscript to be published as INRIA research report 8732; we will also prepare an extract targeting the journal *Argument & Computation*)

ABSTRACT: In this review, we focus on the question of how to structure and retrieve arguments from text. The goal of this review is to identify and place in context a large body of research on argumentation mining and cognate problems. With the forming of a community around the nascent field of argumentation mining, we believe that the time is ripe for a review of the current research problems and approaches. The review covers 160 publications and reports, drawn from work available in English by April 2015.

Jodi Schneider and Stefan Decker. “Socialization of newcomers through quality-control discussions in an open collaboration system”.

(Pending: manuscript, targeting the journal *Computers in Human Behavior*)

ABSTRACT: Socialization of newcomers is an important but difficult task for ad-hoc open collaboration systems that thrive on contributions from members. Volunteered contributions must be subjected to quality controls, yet rejecting contributions can also result in alienating contributors. A prime example is Wikipedia: the online encyclopedia touted as one of crowdsourcing’s greatest successes is facing a crisis in attracting newcomers and transitioning them from readers to editors. One-third of Wikipedia newcomers first contribute by writing a new article and deletion of these articles decreases newcomer retention, while putting a high demand on existing community members for content review. We examine newcomer socialization in quality control conversations within the English-language Wikipedia. In particular, we ask: 1) What are the good and bad outcomes, for newcomers and for the articles undergoing quality control? 2) What are the differences between the rhetoric of newcomers and experienced users in quality control discussions? and 3) What do newcomers need to learn in order to successfully navigate quality control discussions? The discussions around borderline articles provide opportunities for newcomers to receive mentoring, and we argue that, for discussions involving article contributors, one measure of success is whether or not the



newcomer learns community standards, regardless of whether or not the article is subsequently deleted or kept as a part of the encyclopedia. We show that experienced users demonstrate knowledge that newcomers need in order to successfully argue in quality control discussions. This paper has broad implications for online communities that seek contributions from members, and need to check the quality of those contributions.

SWAM2014 Attendees (including **Jodi Schneider**). “From Information Extraction, Information Retrieval and Computational Linguistics to Argumentation Mining”.

(**Pending: manuscript** by SWAM2014 Attendees; my colleagues and I drafted the sections on current approaches and controlled natural language.)

ABSTRACT: How we will get from current fields of information extraction, information retrieval, and computational linguistics, to argumentation mining.

Mathias Brochhausen, Richard Boyce, Paolo Ciccarese, Paul Groth, Yifan Ning, **Jodi Schneider**, Tim Clark. “Curating evidence by modeling scientific claims and facts for a drug-drug interaction use case”.

(**Pending: manuscript**, targeting the ISWC 2015 workshop on Biomedical Data Mining, Modeling, and Semantic Integration)

ABSTRACT: Biomedical claims and facts can be represented in various ways, and several ontologies have been proposed, including the nanopublications ontology, SEE, and the micropublications ontology. The choice of which ontology to use may depend on a variety of factors. In this paper we discuss how integrating multiple evidence representations may be beneficial, using the Drug Interaction Knowledge Base as an example.

Abstracts pending:

Jodi Schneider. “Using argumentation theory and linguistic annotation to analyze and improve online discussions”.

(**Pending: abstract** with outline and bibliography, targeting book chapter on methodology.)

ABSTRACT: Argumentative annotation is a powerful methodology for analyzing online discussions. We describe the uses, applications, and limitations of this methodology.

Jodi Schneider. “Conflict and controversy in discourses around the Wikipedia 'Climate Change' article”.

(**Pending: abstract** with bibliography and sample tools)

ABSTRACT: The Wikipedia article "Climate change" is in the top 250 pages viewed in the English-language encyclopedia with about 150,000 views per month. Over 1600 editors have contributed to the article, which has had over 4000 versions, averaging about once a day since the page was created in April 2002. Using quantitative and qualitative methods, we analyze the conflict and controversy in the article and its creation.

Jodi Schneider. An interface for helping newcomers understand content deletion standards in English Wikipedia.

(**Pending: abstract** with work plan for system improvement and for user evaluation)

ABSTRACT: Interfaces to online discussions have been shown to impact understanding and outcomes of a discussion. Newcomers’ participation in decision-making discussions can be problematic, when community decision standards are not evident. In this paper we describe ongoing work to orient newcomers and to summarize participants’ deliberations about borderline content being considered for deletion from the online encyclopedia Wikipedia.



III – ATTENDED SEMINARS, WORKHOPS, CONFERENCES

- ACM womENCourage Europe, Manchester, UK, February 28- March 1, 2014.
- Argumentation, Dialogue, Persuasion conference, Warsaw, Poland, May 24, 2014.
- Association for Computational Linguistics, First Workshop on Argumentation Mining, Baltimore, Maryland, USA, June 26, 2014.
- American Library Association, Las Vegas, Nevada, USA, June 27-28, 2014.
- Arguing on the Web 2.0, Amsterdam, Netherlands, June 30–July 1, 2014.
- International Society for the Study of Argument, Amsterdam, Netherlands, July 2-4, 2014.
- SICSA Workshop on Argument Mining, Dundee, Scotland, UK July 9-10, 2014.
- Frontiers and Connections between Argumentation Theory and Natural Language Processing, Bertinoro, Italy, July 21-24, 2014.
- International Conference on the Design of Cooperative Systems, Nice, France, May 29-30, 2014
- CrEDIBLE, Federating distributed and heterogeneous biomedical data and knowledge, Sophia Antipolis, France, October 8-10, 2014.
- ABCDE Seminar IV, at the ERCIM 25th Anniversary Celebration, Pisa, Italy, October 23-24, 2014.
- Association for Information Science & Technology meeting, including Doctoral Seminar on Research and Career Development, Seattle, Washington, USA, November 2-4, 2014.
- FORCE2015 Research Communication and e-Scholarship Conference, Oxford, UK, January 12-13, 2015.
- Big Data in a Transdisciplinary Perspective, Herrenhausen Conference series, Hanover, Germany, March 25-27, 2015.

Related to the above, several refereed presentations, not published in proceedings:

Jodi Schneider, Carol Collins, Lisa Hines, John R Horn, and Richard Boyce. “Modeling Arguments in Scientific Papers to Support Pharmacists”. 12th Argumentation, Dialogue, Persuasion conference (ArgDiaP). Warsaw, Poland, May 24, 2014.

Jodi Schneider Some Linked Data Developments from Belgium, Norway, Ireland, and France in “ALA: International Developments in Library Linked Data: Think Globally, Act Globally” [panel]. Las Vegas, Nevada, June 28, 2014

Jodi Schneider, Galen Charlton, Dan Scott, and Richard Urban. Practical Linked Data With Open Source [full-day tutorial], at American Library Association. Las Vegas, Nevada, June 27, 2014

Jodi Schneider, Serena Villata, Elena Cabrio, “Why did they post that argument? Communicative intentions of Web 2.0 arguments”. *Arguing on the Web 2.0* at the 8th ISSA Conference on Argumentation, Amsterdam, The Netherlands, June 30–July 1, 2014.



Jodi Schneider. Evidence Informatics: An Informatics Perspective on Argumentation Mining. SICSA Workshop on Argument Mining: Perspectives from Information Extraction, Information Retrieval and Computational Linguistics. Dundee, Scotland, July 9-10, 2014.

Not attending 2 conferences where I would have presented, due to inadequate travel funding (unable to use ERCIM funds after the end of the fellowship):

Jodi Schneider, “Evidence Informatics.” *1st European Conference on Argumentation: Argumentation and Reasoned Action*, Lisbon, Portugal, June 9-12, 2015

Jodi Schneider, “Fist-clenchingly poor science”: adversarial and collegial social actions in professional public discourse of published scientific papers”. Refereed presentation for *Researching and understanding the language of aggression and conflict*. Panel at 14th International Pragmatics Conference, Antwerp, Belgium, July 26-31, 2015.

IV – RESEARCH EXCHANGE PROGRAMME (REP)

CWI – Netherlands, Information Access research group, Lynda Hardiman, April 20-24

Description: I gave a talk at CWI, on “Synthesizing knowledge from disagreement”, and got very relevant feedback from several members of the department. I was impressed with the variety of research in information, especially their work on cultural heritage. At the CWI library I discovered some research-related books new to me, which I have subsequently bought. I also had discussions with two other ERCIM fellows.

In Amsterdam, I also met with related groups. I spent a half day at Europeana, The Hague, where I gave a talk to the staff on “Packaging ideas: nanopublications in the humanities” and met with the R&D group. I also visited the Amsterdam Medical Centre, where I met with four members of the epidemiology department about my current research in systematic literature reviews.

University of Southampton, Web and Internet Science group, Elena Simperl, April 26-May 1

Description: I had a busy and productive visit, meeting with several faculty and PhD students each day. Topics ranged from citizen science and social machines to argumentation & narrative, provenance, sensemaking, and software sustainability. I also had lunch with a colleague in R&D nearby. Southampton has a significant amount of work that is closely related to my research, so it was valuable to make new connections.

This also allowed me the opportunity to visit the Carole Goble and the Information Management group at the University of Manchester (May 5-7), where I gave a talk on “Synthesizing knowledge from disagreement” and had discussions with Carole’s team and professors in related areas, from biomedical informatics to text mining.