

1. Cover Sheet:

FINAL PERFORMANCE REPORT

**Developing Models for Synchronizing the Interaction among Users, Systems and
Content in Complex Information Spaces**

Bernard J. (Jim) Jansen

329 IST Building
College of Information Sciences and Technology
The Pennsylvania State University
University Park, Pennsylvania 16802
Phone: 814-865-6459
Fax: 814-865-6426
Email: jjansen@acm.org

Award Number: FA9550-06-1-0328
Special Announcement, BAA No. AFOSR 2005-4
Air Force Office of Scientific Research
875 North Randolph St., Suite 325, Room 3112
Arlington, VA 22203-1768

20120918073

2. Objectives:

Specific objectives, unchanged since the proposal submission were:

- a. Methodologies for utilizing the implicit meanings during exchanges as users articulate and reformulate their information need to identify relevant data.
- b. Models that use the implicit feedback, contextual aspects, and cognitive expression of interactions between users and system to refine the user's need.
- c. Approaches to analyze and leverage the linkage between Web documents and queries.
- d. A framework that synchronizes these three separate interaction components (i.e., user-information, user-system, and system-information) to provide the right information in the right formation at the right time to the right set of users.
- e. Methods of evaluating these models for further improvements and adaptation to changing systems, users, and content.

3. Overview of Achievements:

I could not be more pleased with the outcome of this research project in achieving the objectives set forth in the original proposal, along with the research productivity.

As a review, upon receipt of the award, the principal investigator (PI) formed an demographically / intellectually diverse team composed of a tenure track faculty member (the PI), a research assistant with considerable military research experience, four graduate students (one from Information Sciences and Technology, one from Computer Science, one from Industrial and Manufacturing Engineering and one from Electrical Engineering), and one undergraduate student from Information Sciences and Technology. Over the course of the project, some students moved on, and others were added. In total, the project supported: one faculty member (summer support), five graduate students, and seven undergraduate students.

During the latter part of the first year and during the second year, the research team has made significant progress on research goals (a), (b), and (c) as listed in section 2. Objectives. Using massive amounts of log data from major Web search engine companies, we explored methodologies for utilizing the implicit expressions during interactions to articulate the users' needs in order to identify relevant data. Using n-gram approaches, we have begun developing the models of implicit feedback and cognitive expression of interactions to understand how users refine their needs. Finally, we have developed novel and successful approaches to analyze and leverage the linkage between content and expressions of user intent.

The third year was also productive, with the continued progress in achieving the objectives set forth in the original proposal. Building upon the efforts of the first two years, the project progressed to more refined modeling techniques that enabled the inference of actions within an information space, in direct support of

(d) and (e) listed in section 2. Objectives. Specifically, we investigated the range of factors present in information (including behavioral, cognitive, affective, and situational) and which factors foretell future actions. Our research led to a paradigm shift of viewing information not as a static entity, but rather as a dynamic, temporal data stream where information has meaning only in the context of what has come before it and what may come after it. Our aim was to investigate algorithmic methods, building on existing approaches such as binary trees, clustering, n-grams, Markov chains, neural networks, time series analysis, and tensor analysis, which we accomplished.

Overall, we accomplished goals 2a, 2b, 2c, and 2d, although additional work can be done. We made some progress on goal 2e.

4. Research Highlights

Significance to the Field: Determining intent underlying users' request to information technology systems has been a long sought after goal, with substantial work occurring in a variety of fields (i.e., information science, computer human interaction, and management information systems).

Using data mining approaches and temporal analysis techniques (e.g., times series analysis and tensor analysis), we have developed algorithmic formulas that can describe and predict aspect of a user's searching behaviors. We have begun combining this individual aspect with clustering of user queries utilizing the underlying user intent (information, navigational, or transactional) based on quantitatively identified attributes. These modeling techniques permit one to design information technology systems to better support users in complex information systems. These techniques can be used to match user intentions with classified content in order to improve the ability of information technology systems to respond to user goals. With this as a starting point, we would like to explore and develop temporal techniques to model the characteristics and interactions within the data stream.

Developing a predictive model of user actions in complex information spaces has significant advantages and leverages work occurring in a variety of fields (i.e., information science, computer human interaction, and management information systems). If a system can deduce the user intent and decode user behavior, the information system can provide better information to the user. The underlying construct is that if a system can deduce the user's intent, the system or technology can better satisfy the user's goal. We made substantial progress in this area as evidenced by seven journal articles in prestigious outlets such as the Journal of the American Society for Information Science and Technology and, Information Processing & Management, along with six conference proceeding papers in highly regarded conferences such World Wide Web (WWW), ACM Special Interest Group on Information Retrieval (SIGIR), and ACM Special Interest Group on Computer Human Interaction (SIGCHI). There are also conference papers under review.

Relationship to Original Goals: These research results directly support research goals (a), (b), (c), (d), and (e) listed in item 2 Objectives.

Relevance to the Air Force: The objective of this research project is to advance data gathering, information assimilation, and knowledge sharing within complex information spaces by developing models of interactions between (1) searcher/user and system, (2) searcher/user and information, and (3) system and information. The aimed for end results are robust models of human–system–information interactions within complex information contexts from which one can design interfaces, storage structures, retrieval mechanisms, and collaborative sharing workspaces for information and knowledge systems.

Potential Applications: Military plans and operations benefit from heightened situational awareness and the real-time projection of expertise into and out of the battlefield, so this research is a critical and relevant contribution to the mission of the Air Force. By drawing on existing theory, system, and evaluation techniques, the approach outlined by this technical proposal for the development and evaluation of the models is based on sound scientific and technical merit.

The research results will advance and directly contribute to effective implementation of a shift from passive data collections to active, instantaneous, and synchronized exploitation of actionable information.

5. Personnel Supported:

- Bernard J. Jansen (Faculty)
College of Information Sciences and Technology
The Pennsylvania State University, University Park, Pennsylvania 16802
- Mimi Zhang (Graduate Student)
College of Information Sciences and Technology
The Pennsylvania State University, University Park, Pennsylvania 16802
- Ashish Kathuria (Graduate Student)
Department of Electrical Engineering
The Pennsylvania State University, University Park, Pennsylvania 16802
- Chandrika Gopalakrishna (Graduate Student)
Department of Computer Science and Engineering
The Pennsylvania State University, University Park, Pennsylvania 16802
- Ying Zhang (Graduate Student)
Department of Industrial and Manufacturing Engineering
The Pennsylvania State University, University Park, Pennsylvania 16802

- Vijay Mohan (Graduate Student)
Department of Electrical Engineering
The Pennsylvania State University, University Park, Pennsylvania 16802
- Danielle Booth (Undergraduate Student)
College of Information Sciences and Technology
The Pennsylvania State University, University Park, Pennsylvania 16802
- Kate Sobel (Undergraduate Student)
Smeal College of Business
The Pennsylvania State University, University Park, Pennsylvania 16802
- Lauren Solomon (Undergraduate Student)
College of Communication
The Pennsylvania State University, University Park, Pennsylvania 16802
- Arielle Amchin
Smeal College of Business
The Pennsylvania State University, University Park, Pennsylvania 16802
- Peter Smith
College of Information Sciences and Technology
The Pennsylvania State University, University Park, Pennsylvania 16802
- Simone Schuster
Smeal College of Business
The Pennsylvania State University, University Park, Pennsylvania 16802

6. Publications: List peer-reviewed publications submitted and/or accepted during the project period.

Kuthuria, A. and Jansen, B.J. (Under Review) *K-means Clustering to Determine User Intent of Web Queries*. Internet Research.

Mohan, V. and Jansen, B. J. (Under Review) *Predicting Individual Web User Interactions with Time Series Analysis*. ACM Transactions on the Web.

Jansen, B. J., Booth, D. and Smith, B. (2009) *Using the taxonomy of cognitive learning to model online searching*. Information Processing & Management. 45(6), 643-663.

Jansen, B. J., Zhang, M., and Schultz, C. (2009). *Search engine brand and the effect on user perception of searching performance*. Journal of the American Society for Information Sciences and Technology. 60(8), 1572-1595.

Jansen, B. J., Booth, D. L., & Spink, A. (2009). *Patterns of query modification during Web searching*. Journal of the American Society for Information Science and Technology. 60(3), 557-570. 60(7), 1358-1371.

Zhang, Y., Jansen, B. J., Spink, A. (2009) *Identification of factors predicting clickthrough in Web searching using neural network analysis*. Journal of the American Society for Information Science and Technology. 60(3), 557-570.

Zhang, Y., Jansen, B. J., Spink, A. (2009) *Time Series Analysis of a Web Search Engine Transaction Log*, Information Processing & Management. 45(2), 230-245.

Jansen, B. J., Booth, D., and Spink, A. (2008) *Determining the informational, navigational, and transactional intent of Web queries*, Information Processing & Management. 44(3), 1251-1266

Jansen, B. J., Zhang, M., and Spink, A. (2007) *Patterns and transitions of query reformulation during Web searching*, International Journal of Web Information Systems. 3(4), 328-340.

7. Interactions/Transitions:

a. Participation/presentations at meetings, conferences, seminars, etc.

Jansen, B. J., Booth, D. and Spink, A (2009) *Predicting Query Reformulation During Web Searching*. ACM Conference on Computer Human Interaction (CHI2009). p. 3907-3912. Boston, Massachusetts. 4 - 9 April.

Jansen, B. J., Zhang, M., and Schultz, C. (2008) *The Effect of Brand on the Evaluation of IT System Performance*. Proceedings of the Southern Association for Information Systems Conference, Richmond, VA, USA 13-15 March 2008.

Zhang, Y. and Jansen, B. J. (2007) *An Analysis of Searchers' Perceptions of Sponsored and Non-sponsored Links Using Nested Design*, 2007 Annual Meeting of the American Society for Information Science and Technology. Milwaukee, Wisconsin, 18-25 October.

Jansen, B. J., Zhang, M., and Zhang, Y. (2007) *Brand Awareness and the Evaluation of Search Results*, 16th International World Wide Web Conference (WWW2007), p. 1139 – 1140. Banff, Canada. 8-12 May.

Jansen, B. J., Booth, D., and Spink, A. (2007) *Determining the User Intent of Web Search Engine Queries*, 16th International World Wide Web Conference (WWW2007), p. 1149 – 1150. Banff, Canada. 8-12 May.

Jansen, B. J., Zhang, M., and Zhang, Y. (2007) *The Effect of Brand Awareness on the Evaluation of Search Engine Results*, Conference on Human Factors in Computing Systems (SIGCHI), Work-in-Progress, p. 2471 – 2476. San Jose, California. 28 April - 3 May.

b. Consultative and advisory functions to other laboratories and agencies, especially Air Force and other DoD laboratories.

None

c. Technology Assists, Transitions, and Transfers.

None

8. New discoveries, inventions, or patent disclosures.

None

9. Honors/Awards:

- Danielle Booth (Undergraduate Student) from the College of Information Sciences and Technology, The Pennsylvania State University, University Park, Pennsylvania 16802 was awarded the *Undergraduate Research Award for 2007* from the College of Information Sciences and Technology
- Received significant press cover for the publication:

Jansen, B. J., Booth, D., and Spink, A. (2008) *Determining the informational, navigational, and transactional intent of Web queries*, Information Processing & Management. 44(3), 1251-1266. See Penn State press release at: <http://live.psu.edu/story/29879>
- Best Paper Award for:

Jansen, B. J., Zhang, M., and Schultz, C. (2008) *The Effect of Brand on the Evaluation of IT System Performance*. Proceedings of the Southern Association for Information Systems Conference, Richmond, VA, USA 13-15 March 2008.
- Pete Smith (Undergraduate Student) from the College of Information Sciences and Technology, The Pennsylvania State University, University Park, Pennsylvania 16802 was awarded honorable mention the *Undergraduate Research Award for 2009* from the College of Information Sciences and Technology

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

The public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing the burden, to the Department of Defense, Executive Service Directorate (0704-0188). Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.

PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ORGANIZATION.

1. REPORT DATE (DD-MM-YYYY) 02-10-2009		2. REPORT TYPE Final		3. DATES COVERED (From - To) May 2006 - May 2009	
4. TITLE AND SUBTITLE Developing Models for Synchronizing the Interaction among Users, Systems and Content in Complex Information Spaces				5a. CONTRACT NUMBER	
				5b. GRANT NUMBER FA9550-06-1-0328	
				5c. PROGRAM ELEMENT NUMBER	
				5d. PROJECT NUMBER	
				5e. TASK NUMBER	
6. AUTHOR(S) Bernard J. Jansen				5f. WORK UNIT NUMBER	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) 321G IST Building, College of Information Sciences and Technology, The Pennsylvania State University, University Park, Pennsylvania 16802				8. PERFORMING ORGANIZATION REPORT NUMBER	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) Air Force Office of Scientific Research 875 North Randolph St., Suite 325, Room 3112 Arlington, VA 22203-1768				10. SPONSOR/MONITOR'S ACRONYM(S) AFOSR	
				11. SPONSOR/MONITOR'S REPORT NUMBER(S) AFRL-DSR-VA-TR-2012-0556	
12. DISTRIBUTION/AVAILABILITY STATEMENT A - Approve For Public Release.					
13. SUPPLEMENTARY NOTES					
14. ABSTRACT The objective of this research project is to advance data gathering, information assimilation, and knowledge sharing within complex information spaces by developing models of interactions between (1) searcher/user and system, (2) searcher/user and information, and (3) system and information. The end results will be robust models of human-system-information interactions within complex information contexts from which one can design interfaces, storage structures, retrieval mechanisms, and collaborative sharing workspaces for information and knowledge systems. In these complex situations and information spaces, the interactions among searcher, system, and content are more multifaceted. Drawing on general systems theory to understand the interactions among these three components, one could leverage their explicit and implicit exchanges to change the nature of how users view the information space and need. With a vision of the system consisting of interactions, the nature of information gathering changes from a "query -> results -> re-query" model to a multi-faceted paradigm where all components of the system are actively engaged in gathering information, verifying sources, ensuring trust, sharing information, and anticipating information needs.					
15. SUBJECT TERMS information searching, information modeling, information retrieval, information prediction					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Bernard J. Jansen
					19b. TELEPHONE NUMBER (Include area code) 814.865.6459