To appear in the Special Issue for the 25th Anniversary the Knowledge Engineering Review, 2011.

Interactive Knowledge Capture in the New Millennium: How the Semantic Web Changed Everything

Yolanda Gil
USC Information Sciences Institute
gil@isi.edu

Last updated: January 10, 2011

Abstract

The Semantic Web has radically changed the landscape of knowledge acquisition research. It used to be the case that a single user would edit a local knowledge base, that the user would have domain expertise to add to the system, and that the system would have a centralized knowledge base and reasoner. The world surrounding knowledge-rich systems changed drastically with the advent of the Web, and many of the original assumptions were no longer a given. Those assumptions had to be revisited and addressed in combination with new challenges that were put forward. Knowledge-rich systems today are distributed, have many users with different degrees of expertise, and integrate many shared knowledge sources of varying quality. Recent work in interactive knowledge capture includes new and exciting research on collaborative knowledge sharing, collecting knowledge from web volunteers, and capturing knowledge provenance.

1. Introduction

For this special anniversary issue of the *Knowledge Engineering Review*, I prepared a personal perspective on recent research in the area of interactive knowledge capture. My research interest has always been human-computer collaboration, in particular how to assist people in performing complex, knowledge-rich tasks that cannot be fully delegated to a computer program. This raises many research questions: How can people provide a computer with enough knowledge about the task domain and the user's context to provide adequate assistance? How could we make it easier to provide the knowledge needed? How do we know that the computer understands the knowledge that it is given and can make effective use of it? How can a computer be proactive in asking questions about what it does not know and is needed to improve its behavior? All these questions span a number of research areas including knowledge capture (or knowledge acquisition), intelligent user interfaces, knowledge representation and reasoning, and problem solving and decision making, and they lie at the intersection of artificial intelligence, cognitive science, and human-computer interaction.

Assumptions of Early Knowledge Systems	New Challenges in the Semantic Web	Current Research Topics
• One user	• Many, many, MANY users	• Harnessing volunteer contributions
• User must be trained	• No time for training	• Natural tutorial instruction
• One system	• Many distributed subsystems	• Meta-level reasoning
• One knowledge base	• Many knowledge bases	• OWL and semantic web languages
• One reasoner	• Many reasoners	• Semantic web reasoners
• Expert user	• Varying quality and coverage	• Provenance and trust

Figure 1. The early assumptions made in early knowledge systems were challenged by the Semantic Web, bringing about new research topics for the interactive knowledge capture community.

The landscape of this research area has radically changed with the advent of the Web, and of the Semantic Web in particular. We used to assume that there would be a single user editing the knowledge base, that he or she would have domain expertise to add to the system, that he or she would received some training in advance, and that the system would have a centralized knowledge base and reasoner. Today, these assumptions are no longer a given. We have knowledge bases that are distributed, that are edited by groups of users with varying expertise and training, and that contain knowledge of varying quality. These very radical changes have posed new challenges, and at the same time have revived the field and expanded it well beyond the realm of artificial intelligence research.

This paper describes the evolution of the field as a personal journey, describing solely work within my research group and with collaborators. Although it focuses on my work, the journey should be representative of the changes that the field has undergone and the trends in the research emphasis of other researchers. Figure 1 gives an overarching guide to the topics discussed in the paper. We begin with a journey to the early days of knowledge acquisition.

2. Early Work: The Knowledge Acquisition Bottleneck

Our early research investigated knowledge acquisition in the context of a user teaching a computer how to perform the aspects of the task that he or she wanted to delegate. We assumed the system had a knowledge base that the user was modifying and extending. This was dubbed as "the knowledge acquisition bottleneck" in the 1980's, and there were a number of knowledge acquisition tools developed to address it. Each tool was developed for a specific type of problem solving task. An early example was SALT, developed by expert system pioneer John McDermott and colleagues to acquire knowledge for constraint satisfaction problem solvers [Marcus and McDermott 89].

Our research demonstrated that this proliferation of knowledge acquisition tools was unnecessary and that a single general-purpose knowledge acquisition system with an underlying generic problem solver was sufficient, as long as the problem solving method at hand could be represented in the problem solver's language and the knowledge acquisition system could reason about it with respect to the knowledge that needed to be acquired. In

[Gil and Melz 96], we showed that all the questions that SALT was designed to ask from users were automatically generated by our EXPECT knowledge acquisition system, and that additional questions were created by EXPECT to account for gaps in the knowledge base that SALT did not address. This meant that there was no need to develop a different system for each style of reasoning, rather that one single knowledge acquisition system with more explicit representations was sufficient.

Another issue that we investigated is maintaining a knowledge base consistent through many individual changes [Gil and Tallis 97; Tallis and Gil 99]. We found that the more the knowledge acquisition system can reason about different kinds of gaps in its knowledge, the better it can hypothesize how those gaps could be filled and therefore be more proactive and cause less burden to the user. This was reported in [Blythe et al 01], the last paper written about the EXPECT project and one that received a Best Paper award at the 2001 Intelligent User Interfaces conference.

Through this work, we found that we needed a methodology to assess the quality of knowledge acquisition systems, and were surprised there was not one for this area of research. We designed a methodology for user studies that drew from and extended methods from the literature on evaluating programmers, tutoring systems, and user interfaces [Tallis et al 01]. We have used this methodology over the years to evaluate a number of knowledge acquisition systems in a variety of application domains, with users ranging from biologists to military officers to project assistants [Kim and Gil 99; Kim and Gil 00; Kim and Gil 01; Clark et al 01].

Thanks to these user evaluations and experiments, we were able to obtain important quantitative findings. We learned that users were able to complete complex tasks only using the assistance of a knowledge acquisition system, that users enter knowledge much faster, and that knowledge acquisition systems help regardless of the level of programming knowledge or sophistication of the user.

3. From a Single User to Millions of Knowledge Contributors

Unlike our earlier assumption that a single user would be teaching, there are now vast amounts of people on the Web interested or willing to provide knowledge to computers, whether deep domain expertise or personal preferences. Myriads of web users contribute structured content that ranges from simple labels or tags to complex mashups.

Our research in this area focused on how to acquire common knowledge from volunteer contributors. A broad repository of common knowledge would allow knowledge systems to be less brittle. We investigated how to keep volunteers engaged in contributing, and developed techniques to validate the knowledge acquired [Chklovski and Gil 05b]. An interesting aspect of this work is to accommodate natural language statements, which are very easy for users to provide, while constraining them to be in a form that makes their processing possible. We also designed user interfaces that facilitated the collection of specific types of knowledge from users, including process knowledge and argumentation structures [Chklovski and Gil 05a; Chklovski and Gil 05c]. By 2006, our knowledge repository contained more than 700,000 statements concerning common knowledge, becoming one of the top 100 largest OWL documents on the web [Ebiquity 06].

It became clear very quickly that there are very few universally agreed upon truths and that accommodating different views is important in an open knowledge system such as the web. We are currently working on novel frameworks to create knowledge resources that incorporate alternative views within a community of contributors [Vrandečić et al 10].

4. No Training Required

When teaching a computer a complex task, it is important to receive some training ahead of time concerning the abilities and limitations of the system, the methods to test that it has understood correctly, and most of all the use of a language to communicate knowledge. Long periods of training get in the way to doing the teaching and are generally not acceptable to users. In addition, users much prefer to communicate knowledge in a manner that is natural to them.

Some projections argue that 90 million end users without having programming expertise will be interested in creating applications of reasonable complexity such as spreadsheets, web sites, and computer game content by 2012 [Scaffidi et al 05]. Their "programs" will be shared with others and extended or adapted to other problems. The focus of our research in this area is learning procedure descriptions from users through tutorial instruction. An extensive survey of this area revealed that people find it very natural to teach procedures by giving general instructions but that typically give poor instruction. We used descriptions of data analysis workflows built by biologists and shared in a social web site, and developed a knowledge acquisition system that can take such descriptions in natural language and create workflows for the user [Groth and Gil 09a; Groth and Gil 09b].

We are currently developing an approach to learn procedures from human tutorial instruction given in natural language, where the system uses heuristics to handles ambiguities and other imperfections in the instruction [Gil et al 11].

5. Functioning with Limited Knowledge

Our knowledge acquisition systems enabled unsophisticated users to enter knowledge about complex processes in expert-level domains. Alas, although the knowledge was mostly correct, the missing bits did not allow the system to reason appropriately about the task. Perfect knowledge cannot be a requirement. Knowledge systems must be able to operate despite containing varying quality and coverage. The web is full of broken links, which was traditionally considered unacceptable in hypertext circles. Similarly, knowledge systems cannot assume that the knowledge will be consistent and complete and that it can always be subject to such assumptions typically made for reasoning in first-order logic.

We turned to research on human learning to understand the principles and cognitive approaches that humans take when acquiring knowledge incrementally through a lesson [Kim and Gil 02; Gil and Kim 02; Kim and Gil 03]. We studied this from the student perspective as well as from the tutor's perspective regarding the organization of lessons and the expectations on the student's ability to reason with limited knowledge.

In order to provide a computer with the ability to understand what it knows and what knowledge it is missing, we are investigating meta-reasoning capabilities that reason over the contents of a knowledge base and detect gaps and inconsistencies. At any given point, the system should be aware of its knowledge gaps and decide whether they affect its ability to do specific kinds of problem solving. We have developed a framework to keep track of multiple alternative models of a procedure that evolve over time as additional information is provided [Kim and Gil 07; Gil et al 10a; Kim et al 10]. These techniques are used in our work on learning procedures from natural instruction mentioned above [Gil et al 11].

6. Distributed Problem Solving

In contrast with early knowledge-based systems, today both knowledge and reasoning tend to be physically and organizationally distributed on the web. Often the knowledge sources needed for a task reside in distributed subsystems. Reasoning functions may also be distributed, for example provided by third party-services. Knowledge development may also be distributed, formed by assembling knowledge bases (or ontologies) that are separately developed.

Our research is looking at distributed knowledge sources and distributed reasoning in multi-institutional scientific collaborations. Some institutions may provide access to their data as a repository or as a remote service, others may contribute software and tools to process data. Scientists in the collaboration develop applications that retrieve, process, characterize, and store data. Our work focuses on computational workflows to represent declaratively the application software components and the dataflow among them [Gil 06]. We have developed semantic representations of workflows that can be used to reason about the data and the processing steps to automatically generate workflows as well as metadata descriptions of new data products that result from the computations [Gil 09]. Our algorithms to reason about semantic workflows assume that data repositories and software repositories are distributed and managed separately from the workflow system [Gil et al 10b]. This is a major departure from today's planning and reasoning algorithms, which assume that all the pertinent axioms are accessible to a centralized reasoner. Our algorithms call out to the external repositories with well-defined requests, which cause data-specific or component-specific reasoning at the remote sites.

A major component of this research is a collaboration with grid computing researchers to combine workflow reasoning with workflow execution planning [Deelman et al 03; Deelman et al 05; Maechling et al 05; Gil et al 07a; Gil et al 07b; Kurc et al 09]. We have applied these ideas to improve compiler optimization [Hall et al 08], resource-bound computations [Kumar et al 10], and scientific application performance [Woollard et al 08].

Finding relevant components in distributed architectures is a problem for users. We have developed matchers that are given a natural language description of a task by a user and find relevant agents available to automate that task [Gil and Ramachandran 01; Chalupsky et al 01; Gil and Ratnakar 08]. This capability is crucial for the approaches mentioned above to enable users to create and reuse workflows [Gil et al 09] as well as to specify procedures through tutorial instruction [Groth and Gil 09; Gil et al 11].

7. Provenance and Trust

New knowledge is often obtained from open sources on the web, whether web pages, people, mashups, or services. These sources have varying quality and coverage, and may be incomplete and contradict one another. The information is often of unknown origins and there is often no prior history with many of the sources that may be used to assess their reputation.

One focus of this research is how to determine whether to trust a particular piece of information provided by a source. Most existing trust systems are centered around the notion of entity trust, that is, whether or not to trust a given entity [Artz and Gil 07]. We defined *content trust* as a finer-grained model than entity trust in that it assigns trust to pieces of information rather than the overall entity. We identified major factors that affect trust in information sources [Gil and Artz 07]. We also developed a mathematical model and algorithms to derive source reputation from individual trust ratings [Gil and Ratnakar 02a; Gil and Artz 06].

Another aspect of this research is on tracking the provenance of processed information which is synthesized by processing inputs from many individual sources. Representing the provenance of information is key to be able to track what sources were used and how they were combined [Gil and Ratnakar 02a]. We investigated provenance in knowledge bases, where in order to be trusted each axiom must be tied to the source used to create it [Gil and Ratnakar 02b]. Provenance of knowledge base axioms is even more critical when the knowledge base is to be developed and used by a community, as showed our study of the most widely adopted biomedical ontology to date [Bada et al 04]. We also investigated provenance in workflow systems, and augmented existing provenance models that capture execution details with process and data semantics [Kim et al 08]. As decribed above, our semantic workflow representations can be used to reason about metadata properties of data sources as well as new data products as they are processed by the workflow steps, resulting in detailed provenance records that support semantic queries [Kim et al 06].

In recent work, we aim to create a better understanding of provenance requirements, and to foster shared representations and standards for provenance of Web resources [Moreau et al 10; Gil 10].

8. Conclusions

The influence of the Semantic Web on many research areas in artificial intelligence is palpable. This article discussed my views on how it has influenced the area of interactive knowledge capture and pointed to a number of current research topics in the field. Overall, the sheer overwhelming demand from all netizens of a more useful Web creates ample opportunities for increasing the challenges and potential for broader impact of our research in interactive knowledge capture.

Acknowledgements

I would like to thank all my collaborators over the years for their inspiration and contributions to the research described here. I am also thankful for the funding provided for this work by grants from a number of sources, including the Air Force Office of Scientific Research, the Air Force Research Laboratories, the Defense Advanced Projects Agency, the Intelligence Advanced Research Projects Activity, and the National Institutes of Health, and the National Science Foundation. For this article, I gratefully acknowledge support from the US Defense Advanced Research Projects Agency under grant number HR0011-07-C-0060.

References

[Artz and Gil 07] Donovan Artz and Yolanda Gil. "A Survey of Trust in Computer Science and the Semantic Web", Journal of Web Semantics, Volume 5, Issue 2, 2007.

[Bada et al 04] Michael Bada, Robert Stevens, Carole Goble, Yolanda Gil, Michael Ashburner, Judith A. Blake, J. Michael Cherry, Midori Harris, and Suzanna Lewis. "A Short Study on the Success of the Gene Ontology", Journal of Web Semantics, Vol 1 (2), 2004.

[Blythe et al 01] Jim Blythe, Jihie Kim, Surya Ramachandran, and Yolanda Gil. "An Integrated Environment for Knowledge Acquisition." Proceedings of the ACM International Conference on Intelligent User Interfaces (IUI-01), Santa Fe, New Mexico, January 2001. Recipient of the Best Paper Award.

[Chalupsky et al 01] H. Chalupsky, Y. Gil, C. A. Knoblock, K. Lerman, J. Oh, D. V. Pynadath, T. A. Russ, and M. Tambe. "Electric Elves: Applying Agent Technology to Support Human Organizations", Proceedings of the Thirteenth Annual Conference of Innovative Applications of Artificial Intelligence (IAAI-2001), Seattle, WA, August 2001.

[Chklovski and Gil 05a] Tim Chklovski Yolanda Gil. "User Interfaces with Semi-Formal Representations: A Study of Designing Argumentation Structures", Proceedings of the ACM International Conference on Intelligent User Interfaces (IUI-2005), San Diego, CA, January 2005.

[Chklovski and Gil 05b] Tim Chklovski and Yolanda Gil. "An Analysis of Knowledge Collected from Volunteer Contributors", Proceedings of the Twentieth National Conference on Artificial Intelligence (AAAI-05), Pittsburgh, PA, July 9-13, 2005.

[Chklovski and Gil 05c] Tim Chklovski and Yolanda Gil. "Improving the Design of Intelligent Acquisition Interfaces for Collecting World Knowledge from Web Contributors", Proceedings of the Third International Conference on Knowledge Capture (KCAP-05), Banff, Alberta, October 2-5, 2005.

[Clark et al 01] Peter Clark, John Thompson, Ken Barker, Bruce Porter, Vinay Chaudhi, Andres Rodriguez, Jerome Thomere, Sunil Mishra, Yolanda Gil, Pat Hayes, and Thomas Reichherzer. "Knowledge Entry as Graphical Assembly of Components", Proceedings of the First International Conference on Knowledge Capture (K-CAP'01), Victoria, BC, October 2001.

[Deelman et al 03] Ewa Deelman, Jim Blythe, Yolanda Gil, Carl Kesselman, Gaurang Mehta, Karan Vahi, Kent Blackburn, Albert Lazzarini, Adam Arbree, Richard Cavanaugh, and Scott Koranda. "Mapping Abstract Workflows onto Grid Environments", Journal of Grid Computing, Vol. 1, No. 1, 2003.

[Deelman et al 05] Ewa Deelman, Gurmeet Singh, Mei-Hui Su, James Blythe, Yolanda Gil, Carl Kesselman, Gaurang Mehta, Karan Vahi, G. Bruce Berriman, John Good, Anastasia Laity, Joseph C. Jacob, Daniel S. Katz. "Pegasus: A framework for mapping complex scientific workflows onto distributed systems", Scientific Programming, Vol. 13, No. 3, 2005.

[Ebiquity 06] Ebiquity, "Big OWL documents on the Semantic Web." Available from http://ebiquity.umbc.edu/blogger/?page_id=452.

[Gil 06] Yolanda Gil. "Workflow Composition", In Workflows for e-Science, D. Gannon, E. Deelman, M. Shields, I. Taylor (Eds), Springer Verlag, 2006.

[Gil 09] Yolanda Gil. "From Data to Knowledge to Discoveries: Scientific Workflows and Artificial Intelligence." Scientific Programming, Volume 17, Number 3, 2009.

[Gil 10] Yolanda Gil (Ed). "Report of the W3C Provenance Incubator Group," Available from the World Wide Web Consortium at http://www.w3.org/2005/Incubator/prov/wiki/, 2010.

[Gil et al 07a] Yolanda Gil, Varun Ratnakar, Ewa Deelman, Gaurang Mehta, and Jihie Kim. "Wings for Pegasus: Creating Large-Scale Scientific Applications Using Semantic Representations of Computational Workflows", Proceedings of the 19th Annual Conference on Innovative Applications of Artificial Intelligence (IAAI-07), Vancouver, British Columbia, Canada, July 22-26, 2007.

[Gil et al 07b] Yolanda Gil, Ewa Deelman, Mark Ellisman, Thomas Fahringer, Geoffrey Fox, Dennis Gannon, Carole Goble, Miron Livny, Luc Moreau, and Jim Myers. "Examining the Challenges of Scientific Workflows", IEEE Computer, vol. 40, no. 12, pp. 24-32, December, 2007.

[Gil et al 09] Yolanda Gil, Jihie Kim, Gonzalo Florez, Varun Ratnakar, and Pedro A. Gonzalez Calero. "Workflow Matching Using Semantic Metadata." Proceedings of the Fifth International Conference on Knowledge Capture (K-CAP), Redondo Beach, CA, September 1-4, 2009.

[Gil et al 10a] Yolanda Gil, Jihie Kim, and Marc Spraragen. "Principles for Interactive Acquisition and Validation of Workflows." Journal of Experimental and Theoretical Artificial Intelligence, Vol 22, No 2, 2010.

[Gil et al 10b] Yolanda Gil, Pedro A. Gonzalez-Calero, Jihie Kim, Joshua Moody and Varun Ratnakar. "Automatic Generation of Computational Workflows from Workflow Templates Using Distributed Data and Component Catalogs." Journal of Experimental and Theoretical Artificial Intelligence, forthcoming, 2010.

[Gil et al 11]. Yolanda Gil, Varun Ratnakar, and Christian Fritz. "TellMe: Learning Procedures from Tutorial Instruction." Proceedings of the ACM International Conference on Intelligent User Interfaces (IUI-11), Palo Alto, CA, February 13-16, 2011.

[Gil and Artz 06] Yolanda Gil and Donovan Artz. "Towards Content Trust of Web Resources", Proceedings of the Fifteenth International World Wide Web Conference, (WWW-06), Edinburgh, Scotland, May 2006.

[Gil and Artz 07] Yolanda Gil and Donovan Artz. "Towards Content Trust of Web Resources", Journal of Web Semantics, Volume 5, Issue 4, December 2007.

[Gil and Kim 02] Yolanda Gil and Jihie Kim. "Interactive Knowledge Acquisition Tools: A Tutoring Perspective", Proceedings of the 24th Annual Meeting of the Cognitive Science Society (CogSci), Fairfax, VA, August 8-10, 2002.

[Gil and Melz 96] Yolanda Gil and Eric Melz. "Explicit Representations of Problem-Solving Strategies to Support Knowledge Acquisition." Proceedings of the Thirteen National Conference on Artificial Intelligence (AAAI-96), Portland, OR, August 4-8, 1996.

[Gil and Ramachandran 01] Yolanda Gil and Surya Ramachandran. "PHOSPHORUS: A Task-Based Agent Matchmaker", Proceedings of the Fifth International Conference on Autonomous Agents (Agents'01), short paper track. Montreal, Canada, May 28-June 1, 2001.

[Gil and Ratnakar 02a] Yolanda Gil and Varun Ratnakar. "Trusting Information Sources One Citizen at a Time", Proceedings of the First International Semantic Web Conference (ISWC), Sardinia, Italy, June 2002.

[Gil and Ratnakar 02b] Yolanda Gil and Varun Ratnakar. "TRELLIS: An Interactive Tool for Capturing Information Analysis and Decision Making", Proceedings of the 13th International Conference on Knowledge Engineering and Knowledge Management (EKAW), Siguenza, Spain, October 2002.

[Gil and Ratnakar 02c] Yolanda Gil and Varun Ratnakar. "IKRAFT: Interactive Knowledge Representation and Acquisition from Text", Proceedings of the 13th International Conference on Knowledge Engineering and Knowledge Management (EKAW), Siguenza, Spain, October 2002.

[Gil and Ratnakar 08] Yolanda Gil and Varun Ratnakar. "Automating To-Do Lists for Users: Interpretation of To-Dos for Selecting and Tasking Agents." Proceedings of the Twenty-Third Conference of the Association for the Advancement of Artificial Intelligence (AAAI-08), Chicago, IL, July 13-17, 2008.

[Gil and Tallis 97] Yolanda Gil and Marcelo Tallis. "A Script-Based Approach to Modifying Knowledge Bases." Proceedings of the Fourteenth National Conference on Artificial Intelligence (AAAI-97), Providence, RI, July 27-31, 1997.

[Groth et al 11] Paul Groth, Yolanda Gil, James Cheney, and Simon Miles. "Requirements for Provenance on the Web." Submitted for publication, 2011.

[Groth and Gil 09a] Paul Groth and Yolanda Gil. "A Scientific Workflow Construction Command Line." To appear in Proceedings of the ACM International Conference on Intelligent User Interfaces (IUI-09), Sanibel, FL, February 8-11, 2009.

[Groth and Gil 09b] Paul Groth and Yolanda Gil. "Scaffolding Instructions to Learn Procedures from Users." To appear in AAAI Spring Symposium on Agents that Learn from Human Teachers, Stanford, CA, March 2009.

[Hall et al 08] Mary Hall, Yolanda Gil, and Robert Lucas. "Self-Configuring Applications for Heterogeneous Systems: Program Composition and Optimization Using Cognitive Techniques". Proceedings of the IEEE, Special Issue on Edge Computing, February 2008.

[Kim et al 06] Jihie Kim, Yolanda Gil, and Varun Ratnakar. "Semantic Metadata Generation for Large Scientific Workflows", Proceedings of the Fifth International Semantic Web Conference (ISWC-06), Athens, GA, November 5-9, 2006.

[Kim et al 08] Jihie Kim, Ewa Deelman, Yolanda Gil, Gaurang Mehta, Varun Ratnakar. "Provenance Trails in the Wings/Pegasus Workflow System", Concurrency and Computation: Practice and Experience, Special Issue on the First Provenance Challenge, Vol 20, Issue 5, April 2008.

[Kim et al 10] Jihie Kim, Karen Myers, Melinda Gervasio, and Yolanda Gil. "Goal-Directed Meta-Control for Integrated Procedure Learning." To appear in Metareasoning: Thinking about Thinking, Cox, M.T. and Raja, A. (Eds), MIT Press, 2010.

[Kim and Gil 99] Jihie Kim and Yolanda Gil. "Deriving Expectations to Guide Knowledge Base Creation." Proceedings of the Fourteenth National Conference on Artificial Intelligence (AAAI-99), Orlando, FL, July 19-22, 1999.

[Kim and Gil 00] Jihie Kim and Yolanda Gil. "Acquiring Problem-Solving Knowledge from End Users: Putting Interdependency Models to the Test." Proceedings of the Fifteenth National Conference on Artificial Intelligence (AAAI-2000), Austin, TX, July 30-August 3, 2000.

[Kim and Gil 01] Jihie Kim and Yolanda Gil, "Knowledge Analysis on Process Models," Proceedings of the Seventeenth International Joint Conference on Artificial Intelligence (IJCAI-2001), Seattle, WA, August 2001.

[Kim and Gil 02] Jihie Kim and Yolanda Gil. "Deriving Acquisition Principles from Tutoring Principles", Proceedings of the Intelligent Tutoring Systems Conference (ITS), Biarritz, France, June 2002.

[Kim and Gil 03] Jihie Kim and Yolanda Gil. "Proactive Acquisition from Tutoring and Learning Principles", Proceedings of the 11th International Conference on Artificial Intelligence in Education (AI-ED), July 20-24, 2003, Sydney, Australia.

[Kim and Gil 07] Jihie Kim and Yolanda Gil. "Incorporating Tutoring Principles into Interactive Knowledge Acquisition", International Journal of Human-Computer Studies, (65)10, October 2007.

[Kumar et al 10] Vijay Kumar, Tahsin Kurc, Varun Ratnakar, Jihie Kim, Gaurang Mehta, Karan Vahi, Yoonju Lee Nelson, P. Sadayappan, Ewa Deelman, Yolanda Gil, Mary Hall, Joel Saltz. "Parameterized Specification, Configuration, and Execution of Data-Intensive Scientific Workflows." Cluster Computing Journal, Vol 13, 2010.

[Kurc et al 09] Tahsin Kurc, Shannon Hastings, Vijay Kumar, Stephen Langella, Ashish Sharma, Tony Pan, Scott Oster, David Ervin, Justin Permar, Sivaramakrishnan Narayanan, Yolanda Gil, Ewa Deelman, Mary Hall, and Joel Saltz. "High Performance Computing and Grid Computing for Integrative Biomedical Research." Journal of High Performance Computing Applications, 23(3), 2009.

[Marcus and McDermott 89] Sandra Marcus, John P. McDermott. "SALT: A Knowledge Acquisition Language for Propose-and-Revise Systems." Artificial Intelligence, 39(1): 1-37, 1989.

[Maechlin et al 05] P. Maechling, H. Chalupsky, M. Dougherty, E. Deelman, Y. Gil, S. Gullapalli, V. Gupta, C. Kesselman, J. Kim, G. Mehta, B. Mendenhall, T. Russ, G. Singh, M. Spraragen, G. Staples, and K. Vahi. "Simplifying Construction of Complex Workflows for Non-Expert Users of the Southern California Earthquake Center Community Modelling Environment", ACM SIGMOD Record, special issue on Scientific Workflows, Vol 34 (3), September 2005.

[Moreau et al 10] Luc Moreau, Ben Clifford, Juliana Freire, Joe Futrelle, Yolanda Gil, Paul Groth, Natalia Kwasnikowska, Simon Miles, Paolo Missier, Jim Myers, Beth Plale, Yogesh Simmhan, Eric

Stephan, and Jan Van den Bussche. The Open Provenance Model Core Specification (v1.1). Future Generation Computer Systems, July 2010.

[Scaffidi et al 05] Christopher Scaffidi, Mary Shaw, and Brad Myers. "Estimating the Numbers of End Users and End User Programmers." Proceedings of the 2005 IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC 2005), Dallas, TX, September 2005.

[Tallis and Gil 99] Marcelo Tallis and Yolanda Gil. "Designing Scripts to Guide Users in Modifying Knowledge-Based Systems." Proceedings of the Fourteenth National Conference on Artificial Intelligence (AAAI-99), Orlando, FL, July 19-22, 1999.

[Tallis et al 01] Marcelo Tallis, Jihie Kim, and Yolanda Gil. "User Studies of Knowledge Acquisition Tools: Methodology and Lessons Learned". Journal of Experimental and Theoretical AI, Vol 13, pp 359-378, 2001.

[Vrandečić et al 10] Denny Vrandecic, Varun Ratnakar, Markus Krötzsch, and Yolanda Gil "Shortipedia: Aggregating and Curating Semantic Web Data." Third Prize in the Semantic Web Challenge, 2010. Available from http://challenge.semanticweb.org/.

[Woollard et al 08] David Woollard, Nenad Medvidovic, Yolanda Gil, and Chris Mattmann. "Scientific Software as Wokflows: From Discovery to Distribution". IEEE Software, Special Issue on Developing Scientific Software, July/August 2008.