

Toward Meaningful Computing

Computer science can revolutionize the humanities as it has the hard sciences by revealing and contextualizing the meaning hidden in digital libraries.



Current initiatives by Google, Yahoo, and a consortium of European research institutions to digitize the holdings of major research libraries worldwide promise to make the world's knowledge accessible as never before.

Yet in order to completely realize this promise, computer scientists must still develop systems that deal effectively with *meaning*, not just with data and information. This grand research and development challenge motivates our call here to improve collaboration between computer scientists and scholars in the humanities.

Research in the humanities (such as historical lexicography and textual criticism) focuses on the interpretation and elucidation of what is meant by a particular text or artifact in its context of origin, expression, and reception. As with recent computational collaborations with biologists and other physical scientists that have revolutionized these fields and spawned subindustries of computer science (such as bioinformatics and physical simulation), collaboration between computer scientists and humanities scholars promises to improve and deepen everyone's connection to and understanding of the emerging global universe of knowledge.

Computer science researchers and practitioners should jump into this challenge, establishing collaborations with humanities scholars, industry leaders,

and programmers and IT experts. Moreover, research and funding agendas should be realigned to include support for the development of humanities computing at all levels (such as in tools for dealing with highly disparate writing systems or text-navigation systems).

To deal with meaning, one must deal with the context of information in the broadest sense. Although dealing with simple question answering or text retrieval might ignore the distinction between meaning and information, more complex tasks require approaches that are more aware of the context of a given text. Context includes the text's relation to other texts, its authors and editors, where and under what circumstances it was written and published, and even who has read it and what they may have had to say about it. What, for example, might be the implications of enhancing Wikipedia with information linking articles with metadata about their authors and editors, even their readers and reviewers? For one thing, users would have social access to Wikipedia's version histories and thus a more complete picture of the development of knowledge in Wikipedia. The technical hurdles for even this straightforward application are nontrivial, demanding careful user-interface design and algorithmic support.

Computer scientists must involve humanities scholars in such development, bringing to bear their centuries of experience making explicit the invisible but crucial connections among texts, people, and

social environments. The great power of the Web is derived in part from the ubiquitous hyperlinks connecting its millions of pages and sites, so parts of its knowledge structure become explicit. But most of the important connections remain implicit, and many more implicit and inaccessible structures will need to be teased out as the world's libraries come online. Thus, a fundamental task to be addressed by a new humanities/computing collaboration is the identification and explication of implicit links, along with development of effective interaction modes—a new generation of “knowledge browsers.”

Over the next few years, this effort may involve identifying and exploiting connections among simi-

lems of searching and analysis (such as how to represent complex social networks involving individuals with a variety of attributes like age, sex, and economic class) and with different roles in the literature (such as authors, readers, and fictional characters). Nonsocial networks of interest also abound, ranging from the relatively simple (such as scene locations in a play) to the relatively complex (such as thematic units shared in the texts of a particular literature). As we develop such contextually aware search tools, other techniques of computational linguistics, image processing, and social network analysis must also be brought to bear, aiming to find yet more complex and higher-level connections among “idea units” in media.

A fundamental task to be addressed by a new humanities/computing collaboration is the identification and explication of implicit links, along with development of effective interaction modes—a new generation of “knowledge browsers.”

lar significant phrases or structures in millions of different documents or using manually assigned metadata (in a Semantic Web [1] sense) to identify connections among documents. Such markup is indeed important; much effort in humanities computing today is devoted to developing XML markup schemes for the structural and semantic properties of texts. For example, the Text Encoding Initiative (www.tei-c.org) is developing many encoded data sets, ranging from the works of individual authors to massive collections of national, historical, and cultural literatures [2]. However, many of the formal assumptions underlying XML markup (such as the idea that documents are essentially trees) are insufficient for representing truly useful markup (many types of textual structures may overlap). Hence, future advances toward making the meaning in digital literature globally accessible depend on the development of richer structural markup schemes that also allow for efficient search and retrieval.

The complex web of relationships within and among texts and their contexts leads to novel prob-

The development of such tools for contextually aware searching and information processing may revolutionize humanities scholarship. Scholars will be able to interact with texts in new ways, particularly as linguistic, visual, and statistical processing give them new ways of reading, visualizing, and understanding them.


The converse—using computational methods to develop new modes of artistic and scholarly expression—is also key. Hypertext, blogging, and new forms of digital media and gaming are all being applied to humanities research and teaching. These emerging networked interactive media represent only the tip of the iceberg.

The emerging global digital library will thus require computer scientists to develop new query interfaces that represent the textual and contextual elements of human meaning. Occurrences of terms in documents may be filtered by rhetorical or linguistic context (such as when a word is used in syntactically, semantically, or thematically salient positions). Visualization tools must include relevant documents

on timelines, geographic representation, author characteristics, or degree of connection to other documents. Meanwhile, the contingency and ambiguity of meaning in human texts pose problems of how to present the content and context of text to the user in a comprehensible and manipulable fashion.

In our expanding information universe, computer scientists must ensure that access enhances and enriches everyone's meaningful experience with information, rather than dehumanizes it by possibly omitting its context. They must join hands with humanists, whose concern with explicating meaning in all its complexity crosses disciplinary boundaries, from literature and philosophy to history and music.

Such collaboration begins with humanist organizations, including the Association for Computing in the Humanities (www.ach.org) and the Association for Literary and Linguistic Computing (www.allc.org). The next international Digital Humanities conference (www.allc-ach2006.colloques.paris-sorbonne.fr) will be held in Paris in July where computer scientists will be able to draw inspiration from and suggest solutions to problems in humanities computing.

This work is sure to yield valuable results in information retrieval, computer graphics, natural language processing, data mining, information visualization, and human-computer interfaces. The result will be that context is added to ubiquitous information access, making it both meaningful and full of meaning. 

REFERENCES

1. Berners-Lee, B., Hendler, J., and Lassila, O. The Semantic Web. *Scientific American* (May 2001).
2. Ide, N. and Sperberg-McQueen, C. The TEI: History, goals, and future. *Computers and the Humanities* 29, 1 (1995), 5–15.

SHLOMO ARGAMON (argamon@iit.edu) is an associate professor in the Computer Science Department of the Illinois Institute of Technology, Chicago.

MARK OLSEN (mark@barkov.uchicago.edu) is assistant director of the Project for American and French Research on the Treasury of the French Language at the University of Chicago.

STAY ON TOP OF ACM NEWS WITH MEMBERNET

IN THE CURRENT ISSUE OF MEMBERNET:

- ACM lends leadership to top research firms.
- The ACM Fellows for 2005 are announced.
- New professional courses from ThomsonNETg in the Professional Development Centre.
- Previews of the International Collegiate Programming Contest and SIGCSE educational symposium.

And much more!

All online, in MemberNet:
www.acm.org/membernet.