

The following resources related to this article are available online at www.sciencemag.org (this information is current as of August 13, 2009):

Updated information and services, including high-resolution figures, can be found in the online version of this article at:

<http://www.sciencemag.org/cgi/content/full/325/5942/828>

This article **cites 28 articles**, 4 of which can be accessed for free:

<http://www.sciencemag.org/cgi/content/full/325/5942/828#otherarticles>

This article appears in the following **subject collections**:

Scientific Community

http://www.sciencemag.org/cgi/collection/sci_commun

Information about obtaining **reprints** of this article or about obtaining **permission to reproduce this article** in whole or in part can be found at:

<http://www.sciencemag.org/about/permissions.dtl>

Strategic Reading, Ontologies, and the Future of Scientific Publishing

Allen H. Renear* and Carole L. Palmer

The revolution in scientific publishing that has been promised since the 1980s is about to take place. Scientists have always read strategically, working with many articles simultaneously to search, filter, scan, link, annotate, and analyze fragments of content. An observed recent increase in strategic reading in the online environment will soon be further intensified by two current trends: (i) the widespread use of digital indexing, retrieval, and navigation resources and (ii) the emergence within many scientific disciplines of interoperable ontologies. Accelerated and enhanced by reading tools that take advantage of ontologies, reading practices will become even more rapid and indirect, transforming the ways in which scientists engage the literature and shaping the evolution of scientific publishing.

The 1980s abounded in descriptions of a coming new world of scholarly communication, predicting functionality that we knew was possible and would soon be technologically feasible. This imagined world, which was never fully realized, predicted advanced navigation; discipline-specific intelligent tools for searching, browsing, and analysis; reader-initiated hypertext linking; “live” data-driven diagrams; computationally available information objects; searchable indexed annotations; thorough-going interoperability; and so on. Substantial improvements in hardware and software and an infrastructure of networked communications now make this anticipated functionality possible. Lying at the heart of the changes taking place is an escalation of strategic reading practices.

Scientists have always read strategically, working with many articles simultaneously to search, filter, compare, arrange, link, annotate, and analyze fragments of content. Now, however, two important trends are interacting to support and intensify the effectiveness of these practices. The first is the wide-scale use by scientists of digital indexing, retrieval, and navigation resources (such as PubMed, Web of Science, the ACM Digital Library, NASA’s Astrophysics Data System, CiteSeer, Scopus, and Google Scholar) to exploit large quantities of relevant information without reading individual articles. The second is the emergence within many scientific disciplines of ontologies for representing and linking scientific data. This convergence of digital resources and data-linking ontologies will result in even more rapid and indirect use of the literature, supported not only by text mining (1) and literature-based

discovery applications (2), but by “ontology-aware” strategic reading tools as well.

Why Will the Revolution Happen Now?

When it was launched in 1992, the *Online Journal of Current Clinical Trials*, jointly designed by the American Association for the Advancement of Science and the OCLC Online Computer Library Center, was seen by some as the beginning of the long-awaited world of advanced digital publishing. However, the journal failed to flourish, and the new world did not materialize. In retrospect, we can see that in the early 1990s, none of the basic conditions required for an advanced scientific publishing system existed. Not only was the basic technology and infrastructure inadequate, but the entire publishing system also would have required extensive coordinated changes.

Although there was no revolution, an important transformation did take place in the 1990s. In 1993, very few scientific, technical, and medical (STM) journals had an electronic version, and yet by 2003, virtually all of them did. For the daily work routines of most scientists, that new format had already become more important than print. The system of digital publishing that emerged from 1993 to 2003 was impressive in some respects, but was still largely another case of new technology compromised by imitation of the old.

The reasons a more radical change failed to occur are understandable in retrospect, and they also suggest why we are now on the cusp of a larger change. None of the developments during this period required costly or uncer-

tain changes in workflow and production processes, software tools, user behavior, or business models. STM publishers were already creating Adobe PostScript files for print production. These could be automatically converted to the Adobe page description language format (PDF), which was suitable for distribution over the existing Internet and could be browsed with existing free software applications. Users, in turn, were presented with a printlike experience that was at once familiar and yet had additional advantages, including Internet delivery, digital storage, full-text searching, and local printing, all of which was easily realized with existing technologies. Hence, as its value became apparent, PDF-based digital STM publishing emerged relatively quickly, with few changes in production or existing infrastructure.

Since 1992, processor speeds, memory, storage, and bandwidth to the desktop have undergone enormous improvements, as have connectivity and costs. Standard protocols for network communication have been adopted, new software tools and software engineering strategies have emerged, and there is now a supporting infrastructure of information professions and institutions. The pervasive use of the World Wide Web via intuitive Web browsers is an especially visible change, and the widespread use of Extensible Markup Language (XML), with its associated standards and technologies, provides a foundational framework for storing, processing, and presenting information on the Web. In addition, an important recent development is the convergence within the STM publishing community on a single XML schema for the representation of scientific articles: the National Library of Medicine (NLM)’s Journal Archiving and Interchange Tag Suite (3).

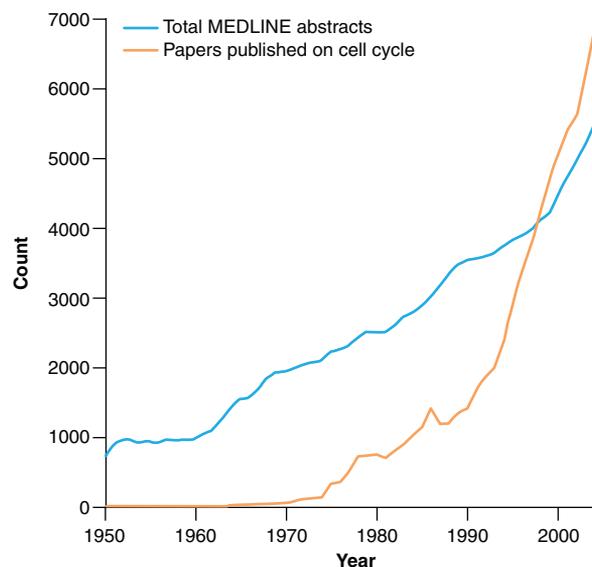


Fig. 1. Increase in number of papers published each year in biomedicine and in one specialized topic, the cell cycle. [Adapted with permission from MacMillan Publishers, *Nature* (5), copyright 2006]

Center for Informatics Research in Science and Scholarship, Graduate School of Library and Information Science, University of Illinois at Urbana-Champaign, Champaign, IL 61820, USA.

*To whom correspondence should be addressed. E-mail: renear@illinois.edu

The driving force for change remains the same: the growing quantity and complexity of information in combination with limited time for reading. But in some disciplines, we seem to be past the point where any further specialization of research focus or elaboration of collaborative relationships are effective (4, 5) (Fig. 1). Just as the increased quantity of information and general intensity of scientific activity is reaching the point where it cannot be sustained with current practices, technology and user behavior are making new practices feasible, and research scenarios that a decade ago were utopian are now widely anticipated by practicing scientists. P. Bourne, a Public Library of Science journal editor, offers this vision of the near future

the scientific literature will seamlessly provide annotation of records in the biological databases. Imagine reading a description of an active site of a biological molecule in a paper, being able to access immediately the atomic coordinates specifically for that active site, and then using a tool to explore the intricate set of hydrogen-bonding interactions described in the paper... Alternatively, if you are starting with the data ... viewing the chromosome location of a human single-nucleotide polymorphism associated with a neurological disorder, ... immediately access a variety of papers ranked in order of relevance to your profile ... pinpointing the reference to the single-nucleotide polymorphism in the full-text article (6), p. 179.

This sort of information gathering goes well beyond conventional digital publishing and reveals why the current state of affairs has failed to meet some expectations. As chemists P. Murray-Rust and H. S. Rzepa remarked in 2004, "The current transition to [PDF-based] e-journals seems to be welcomed by many—but not us ... a cultural change in our approach to information is needed" (7).

How Are Scientists Working with the Literature?

Scientists have always strived to avoid unnecessary reading. Like all researchers, they use indexing and citations as indicators of relevance, abstracts and literature reviews as surrogates for full papers, and social networks of colleagues and graduate students as personal alerting services. The aim is to move rapidly through the literature to assess and exploit content with as little actual reading as possible. As indexing, recommending, and navigation has become more sophisticated in the online environment, these strategic reading practices have intensified.

Now, as scientists search and browse, they are making queries and selecting information in much tighter iterations and with many different kinds of objectives in mind, almost as if they were playing a fast-paced video game. They sweep

through resources, changing search strings, chaining references backward and citations forward, dodging integrator and publisher sites to find open-access copies, continually working to reduce the number of clicks required for access. By note-taking or cutting and pasting, scientists often extract and accumulate bits of specific information, such as findings, equations, protocols, and data. In this process, rapid judgments are made—such as assessments of relevance, impact, and quality—while search queries are being formulated and refined. (Fig. 3). The goal often seems to be undifferentiated assimilation of information about a domain or a problem at hand, and the online experience may be highly valuable, even though no clear aim is met and no articles to read are located. In a compelling analogy, Nicholas *et al.* (8) describe a "slightly irritated" father watching his young daughter flick from channel to channel while watching television

[the] father asks ... why she cannot make up her mind and she answers that she is not attempting to make up her mind but is watching all the channels. ... gathering information horizontally, not vertically (8), p. 40.

And they conclude

Now we see what the migration from traditional to electronic sources has meant in information seeking terms. We are all bouncers and flickers, and the success of Google is a testament to that, with its marvelous ability to enhance and amplify this flicking and bouncing (like a really good remote)... In the past, information seeking was seen to be the first step to creating knowledge. Now ... it is a continuous process (8), pp. 41–42.

Just as the aim of channel surfing is not to find a program to watch, the goal of literature surfing, is not to find an article to read, but rather to find, assess, and exploit a range of information by scanning portions of many articles. This behavior is common among scientists (9).

Longitudinal studies of e-journal use confirm that scientists are indeed "reading" more papers at a faster pace (10). That is, the total time spent reading journal articles has risen only a little, whereas the number of journal articles read per year has gone up much faster and appears to be growing still. The number of articles read (as distinguished from those merely browsed) by scientists was ~50% higher in 2005 than in the

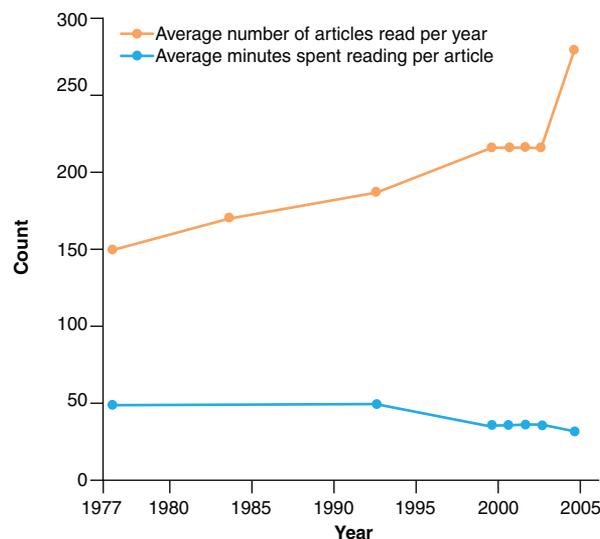


Fig. 2. Increase in the number of papers read by scientists per year and decrease in minutes spent reading each paper, trends based on a series of survey studies conducted by Tenopir *et al.* between 1977 and 2005 (10, 34, 35).

mid-1990s. Furthermore, though the average reading time per article did not change much from 1977 to the mid-1990s (48 versus 47 min), it started falling in the mid-1990s and is now just over 30 min per article (Fig. 2). At the same time, identifying papers by searching online increased more than fourfold between 1977 and 2005. These changes in journal use are far greater in STM disciplines than the averages over all disciplines, suggesting that as work with the literature has moved online, scientists are scanning more and reading less.

Early digital library research also showed how scientists scan individual printed journal articles to identify key components—such as tables of contents, references, figures, formatted lists, equations, and scientific names—for quick review and absorption of information (11, 12). More recent studies of the research process have emphasized the varied ways in which scientists work with information (13, 14). The literature is scanned not only to position new findings in cognate fields and learn about collaborators' domains, but also to monitor the progress of peers and competitors. Information is collated to compare measurement and instrumentation details; it is also used to compile personal collections in evolving areas of interest and to extract the facts and evidence needed to build databases. These are all aspects of strategic reading, a robust, well-entrenched behavior that is vastly more efficient in the digital realm and is thus a promising target for digital support.

How Is Scientific Information Being Represented?

Structured terminologies for representing scientific data, along with standard XML-based techniques for defining and using these terminologies, are forming the basis for new types of scientific publishing. Although computer-processible scientific terminologies range from simple standardized

vocabularies to sophisticated formal systems with logical axioms, we have called all of them ontologies.

Ontologies are particularly prominent in the biological sciences (15, 16). One example of rapid adoption is the Gene Ontology (GO) (17), which started in 1998 to support the annotation of genes and gene products and is now very widely used, containing more than 25,000 terms and 3.3 million annotations. Although many biological ontologies were originally developed independently, the need for interoperability has driven collaboration, a good example being the Open Biomedical Ontologies (OBO), which currently has 54 participating projects (18), including Microarray Gene Expression Data (MGED), BioPAX, for biological pathways data, and Foundational Model of Anatomy (FMA).

Although the size, complexity, and logical design of scientific ontologies may vary, a partial description of GO, drawing on examples from the GO introductory material, will illustrate some of their general features (19, 20). GO consists of three separate ontologies: (i) molecular function, (ii) biological process, and (iii) cellular component. Within each of these, terms are uniquely identified, defined, and related in a network of “is a” relationships (e.g., a nuclear chromosome is a chromosome). GO also contains the relationship “part of” (e.g., periplasmic flagellum is part of periplasmic space), and recently, the relationship “regulates” and subtype relationships “positively regulates” and “negatively regulates” were added. These relationships have logical features; for instance, “is a” and “part

of” are transitive (e.g., if X is part of Y and Y is part of Z, then X is part of Z). It is easy to see not only how the controlled vocabulary of a shared ontology can facilitate the integration of data from multiple sources, but also how relationships such as “is a”, “part of”, and “regulates” can support other information management tasks as well, including information retrieval and text mining, error checking, and automated inferencing.

Neither controlled vocabularies nor even logic-based ontologies are entirely new, although the enormous increase in the amount and complexity of biological data makes such organizational strategies increasingly urgent. Now, however, we can make ontologies and their applications computationally available and interoperable through well-supported standards associated with the Internet and World Wide Web.

In 1998, as work began on GO, the World Wide Web Consortium (W3C) released XML (21), a metalanguage for defining markup languages (22) for representing information on the World Wide Web. Originally designed for document-oriented languages, XML was soon used for other kinds of information as well. XML languages are defined by a computer-readable schema, which specifies, among other things, the terms of the markup language and the ways those terms can be arranged in valid documents. XML organizes information as a hierarchical structure (an “ordered tree”) of labeled nodes and attribute/value pairs and represents that structure in a linear format readable by both humans and computers. Software can read data in this format and construct the correct tree

structure, even without the schema that defines the language (this is a virtue of XML). If a schema is available, additional processing is possible, such as verifying that the data are complete and correctly organized; a schema can also configure editing tools so that human coders are only offered legal coding options, making coding easier and syntax errors impossible. In just 10 years, XML and related supporting software and standards have come to dominate information representation in networked environments—all popular Web browsers support XML, and most major database systems import and export XML-formatted data.

Although using XML to declare and apply a terminological vocabulary improves interoperability and access to software applications, it does have some limitations. XML schemas specify syntax, not semantics (23). An XML schema does not itself indicate how to interpret portions of a particular XML tree structure in terms of scientific assertions, nor is it, alone, suitable for defining logical relationships among terms. That information must be recorded in the natural language documentation for the schema, but then it is unavailable for computer processing. To address this problem, the semantic Web languages Resource Description Framework (RDF), Resource Description Framework Schema (RDFS), Web Ontology Language (OWL), and Semantic Web Rule Language (SWRL) were developed (24). These are computer-processible knowledge representation languages that provide a standard technique for defining ontologies and expressing assertions that use terms from those ontologies. Although technically independent of any particular computer-encoding format, RDFS and OWL each have a standard XML syntax that is now well-supported by software applications and widely used for ontology representation.

Today, an emerging infrastructure of education, research, conferences, organizations, and software tools is sustaining the development and adoption of scientific ontologies and providing opportunities for coordination to improve interoperability and share best practices. Particularly important for biology are the National Center for Biomedical Ontology, OBO, and the International Society for Biocuration, as well as more broadly defined organizations such as the National Center for Biotechnology Information and the European Bioinformatics Institute. One notable software application for ontology development is the widely used and well-supported Protégé ontology editor.

How Can Ontologies Help Scientific Publishing?

Originally motivated by the need for data integration, scientific ontologies are now being explored for STM publishing to support information retrieval and text mining, with applications for hypothesis generation and knowledge discovery well underway. Nevertheless, reading-like engagement with scientific articles is not likely to disappear entirely: The natural language prose of scientific articles provides too much valuable

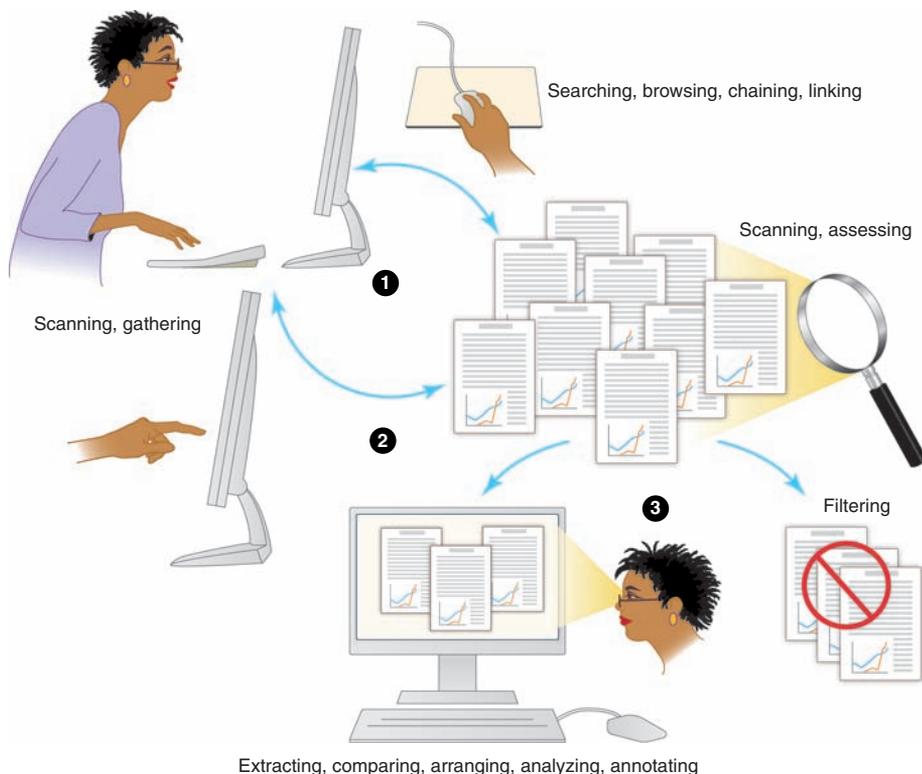


Fig. 3. Current work with digital resources.

Textpresso for *C. elegans*

About Textpresso
Categories/Synonyms
Copyright
Document Finder
Downloads
Feedback
Home
Query Language
Search
User Guide

1516 matches found in 74 documents. Search time: 1.148 seconds.

Global links/files: [all results in endnote](#) [all results in print version](#) [all results in xml](#)

Score: 178.00

Keywords ?

lin-36

Exact match Case sensitive Search synonyms ?

Categories ?

List >

biological process (GO) (all) Reset

Select category 2 from list above Reset

Select category 3 from list above Reset

Select category 4 from list above Reset

Advanced Search Options : on off [location (abstract, full text), sorting]

Title: The *C. elegans* gene *lin-36* acts cell autonomously in the *lin-35* Rb pathway .

Authors: Thomas JH Horvitz HR

Journal: Development

Year: 1999

Doc ID: WBPaper0003632

Bibliographic Information

Abstract

Matching Sentences

Match: These observations indicate that whereas activation of the inductive signal transduction cascade by the gonadal ligand is not necessary for the adoption of vulval fates in the absence of negative regulation by *lin-36* and the other *synMuv* genes, the presence of the intracellular signal transduction cascade nonetheless is necessary. [Field: body, subscore: 8.00]

Match: Although the penetrance of the *Muv* phenotype in animals that have lost *lin-36* is essentially the same as that in animals that have lost *lin-36* activity meiotically (69% of phenotype in animals that have lost *lin-36* activity in ABpl or ABpr is substantially lower (2 / 10, excluding double losses). [Field: body, subscore: 8.00]

Match: However, in the absence of both *synMuv* gene-mediated negative regulation and proper subcellular localization of LET-23 receptor, the P (3-8) . p cells variably adopt either vulval or nonvulval fates, despite the presence of the inductive signal. Laser ablation experiments showed that the P (3-8) . p cell fates seen in animals lacking the stimulatory regulatory pathway mediated by *lin-2*, *lin-7*, *lin-10* and the negative regulatory pathway mediated by the 3458 JH Thomas, and H R Horvitz *synMuv* genes. [Field: body, subscore: 7.00]

Match: A *lin-36* :: GFP reporter is expressed in the nuclei of Pn . p cells We determined the expression pattern of a *lin-36* :: GFP reporter construct containing 1 kb of sequence 5 to the ATG start site, the entire *lin-36* open reading frame, all *lin-36* introns and the GFP gene fused in-frame to the last codon of *lin-36*. [Field: body, subscore: 7.00]

Match: Antagonism of the inductive signaling pathway by the Rb-mediated class B *synMuv* pathway The class B *synMuv* genes *lin-35* and *lin-53* encode proteins *lin-35* acts cell autonomously 3457 similar to the products of the mammalian retinoblastoma (Rb) tumor suppressor gene and the Rb-binding protein RbAp48, respectively (Lu and Horvitz, 1998). [Field: body, subscore: 6.00]

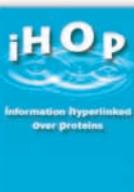
Match: DISCUSSION *lin-36* encodes a novel protein and acts cell autonomously Our results suggest that *lin-36* encodes a novel protein that is required within the P (3-8) . p cells to negatively regulate vulval development. [Field: body, subscore: 6.00]

Match: Key words : *lin-36* , Signal transduction , Vulval development , Redundancy , *C. elegans* INTRODUCTION [Field: body, subscore: 6.00]

Match: We demonstrate that *lin-36* functions in and is expressed in the vulval precursor cells, establishing that the *lin-36* pathway is involved in intercellular signaling. [Field: abstract, subscore: 5.00]

Match: We demonstrate that *lin-36* functions in and is expressed in the vulval precursor cells, establishing that the *lin-36* pathway is involved in intercellular signaling. [Field: body, subscore: 5.00]

Match: We also report that the *lin-36* pathway and / or another pathway that is functionally redundant with the *lin-36* pathway antagonize a ligand-independent activity of the receptor tyrosine kinase / Ras vulval induction pathway. [Field: abstract, subscore: 4.00]



Information Hyperlinked Over Proteins

Search for a gene synonym or accession number... (Click here for an example: SNF1)

SNF1

all fields in All organisms

all fields

synonyms

any accession number

NCBI Gene

UniProt

Google

Symbol	Name	Synonyms	Organism
SNF1	AMP-activated serine/threonine protein kinase found in a complex containing Snf4p and members of the Sip1p/Sip2p/Gal83p family; required for transcription of glucose-repressed genes...	Carbon catabolite-derepressing protein kinase, CAT1, CCR1, D8035.20, GLC2, HAF3, PAS14, YDR477W	Saccharomyces cerevisiae

WikiGenes [edit this page](#) new

UniProt [P06782](#)

IntAct [P06782](#)

PDB Structure [3FAM, 2QLV](#)

NCBI Gene [852088](#)

NCBI RefSeq [NP_010785](#)

NCBI UniGene [852088](#)

NCBI Accession [AAA35058, AAB64904](#)

Homologues of SNF1 ...

Definitions for SNF1 ...

Most recent information for SNF1 ...

Enhanced PubMed/Google query ...

Sentences in this view contain interactions of SNF1 - Interaction Information is available whenever you see this symbol - Read more.

For a summary overview of the information in this page [click here](#). new

We show that [SNF4](#) binds to the [SNF1](#) regulatory domain in low glucose, whereas in high glucose the regulatory domain binds to the kinase domain of [SNF1](#) itself. [1996]

Yeast [Snf4 \[7\]](#) PK-related protein kinases (SnRKs) controlling glucose and stress signaling in eukaryotes is reduced in a sip1delta sip2delta gal83delta triple mutant. [1997]

We first show that [SNF4](#) - "Protein kinase activator found in a complex containing Snf1p and members of ..." (Saccharomyces cerevisiae) ([SNF1](#)) (encoding a protein kinase) and [CAT3](#) ([SNF4](#)) (probably encoding a subunit of ...)

This gene activates [Snf1](#) ([Snf1](#))

The [SNF4](#) -beta-galactosidase protein colimmunoprecipitated with the [SNF1](#) protein kinase, thus providing evidence for the physical association of the two proteins. [1997]

Increased [SNF1](#) gene dosage partially compensates for a mutation in [SNF4](#), and the [SNF4](#) function is required for maximal [SNF1](#) protein kinase activity in vitro. [1989]

MeSH-Term: - Click for options [the Snf4 \(gamma\) subunit in regulating SNF1 protein kinase in response to glucose availability in Saccharomyces cerevisiae. \[2008\]](#)

In two-hybrid assays, one [SNF4](#) mutation enhances the interaction of [SNF1](#) with [Snf4](#). [1997]

Chem. Compound: - Click for options

High impact journal.

Add this sentence to your gene model.

Evidence from large-scale screens for interactions between SNF1 and... -SNF4 (from: TAP & HMS & IntAct)

Fig. 4. Two examples of ontology-aware text mining/retrieval systems that support strategic reading. (Top) Textpresso (www.textpresso.org/) and (Bottom) iHOP (www.ihop-net.org).

nuance and context to be treated only as data (25). Scientists may have moved well beyond traditional reading, but they still remain engaged with the narrative of scientific articles and need tools to help them read, and not only mine, that narrative.

The integration of ontologies into the scientific literature has been recommended by leading scientists (26–28), and the current generation of ontology-based text mining and retrieval tools in the biomedical sciences is already taking advantage of natural language processing and databases of annotations (5, 29, 30). One example is Textpresso, an ontology-based mining and retrieval system that works with prepared collections of articles, split into sentences and annotated with terms from 33 ontology categories, three of which correspond to the GO ontologies (31). Results screens present a ranked list of sentences within a ranked list of articles, with term highlighting, and links to articles and external databases (Fig. 4, top). Reading the sentences of an article in relevance order rather than narrative order is an example of strategic reading within an article. An example of strategic reading across a collection is provided by Information Hyperlinked over Proteins (iHOP), which uses genes and proteins to create a network of sentences and abstracts for searching and navigating MEDLINE abstracts (32). The iHOP database processes abstract sentences using National Center for Biotechnology Information taxonomy identifiers and the Medical Subject Headings (MeSH) thesaurus and supplies pages of configurable results, in ranked lists of sentences retrieved from many abstracts (Fig. 4, bottom).

Unlike similar explorations in the 1980s and 1990s, these are not computer science experiments or pilot projects requiring substantial investment and large upfront changes in infrastructure and practices to scale them up for general use. These are projects that are already producing practical and widely used tools.

How Do We Support and Shape These Changes?

The infrastructures and services to support strategic reading practices will no doubt be promoted by open access and alternative publishing models, which are already being widely discussed in the academic community. However, research on information behavior and the use of ontologies is also needed.

Traditional approaches to evaluating information systems, such as precision, recall, and satisfaction measures, offer limited guidance for further development of strategic reading technologies. Finer-grained methods that analyze what scientists actually do and value are required if we want to understand the nearly subconscious tactics that govern second-by-second interactions with the literature and the nuances of intention and use. We know, for instance, that scientists often have trouble locating very problem-specific information (on methods and protocols, for instance) and that the occasional exploration of

results from another discipline can have considerable impact on progress or the direction of research. These are the kinds of information behaviors that we need to understand more fully to design tools that go beyond search and retrieval to support creative strategic reading.

For ontology-aware reading tools to function well, terminological annotations must be included in, or mapped to, the XML encoding of articles during the publishing production process, to connect names and phrases in narrative text with appropriate standard terminology. The emergence of the NLM schema as a standard XML encoding for scientific articles provides a promising shared context for terminological annotation; however, we also need specific strategies that are economically sustainable within the current context of STM publishing workflows, as well as remedies for “legacy data,” the articles already published and stored in repositories. To exploit terminological annotations across the Internet, reading tools will have to operate in real time to take advantage of the ontologies that define and relate terms and connect terms with relevant databases with the use of “service-oriented architectures” (33). Finally, the development of ontology languages with additional expressive power is needed, as well as continued support for evolving, coordinating, and harmonizing ontologies.

How Will Scientists Work with the Literature in 2019?

Scientists will still read narrative prose, even as text mining and automated processing become common; however, these reading practices will become increasingly strategic, supported by enhanced literature and ontology-aware tools. As part of the publishing workflow, scientific terminology will be indexed routinely against rich ontologies. More importantly, formalized assertions, perhaps maintained in specialized “structured abstracts” (27), will provide indexing and browsing tools with computational access to causal and ontological relationships. Hypertext linking will be extensive, generated both automatically and by readers providing commentary on blogs and through shared annotation databases. At the same time, more tools for enhanced searching, scanning, and analyzing will appear and exploit the increasingly rich layer of indexing, linking, and annotation information.

There are no technical obstacles to this trajectory, and it is already under way. The changes, as always, will be incremental: Scientists, who today already make extensive use of existing indexing and retrieval services, will encounter a steady stream of new enhancements and adopt those that allow rapid and productive engagement with the literature. The new functionality will sometimes be provided as part of the application interface (new features in PubMed, for instance) or as shared external tools that users can add to their Web browsers. These developments chart a middle course between the already obsolete activity of finding an article to read on the

one hand, and the narrower objectives of text mining on the other, responding directly to the entrenched necessity and value of strategic reading in the daily work of today’s scientists.

References and Notes

1. I. Spasic, S. Ananiadou, J. McNaught, A. Kumar, *Brief. Bioinform.* **6**, 239 (2005).
2. D. R. Swanson, N. R. Smalheiser, *Artif. Intell.* **91**, 183 (1997).
3. National Library of Medicine, <http://dtd.nlm.nih.gov/>.
4. B. Mons, *BMC Bioinform.* **6**, 142 (2005).
5. L. J. Jensen, J. Saric, P. Bork, *Nat. Rev. Genet.* **7**, 119 (2006).
6. P. Bourne, *PLoS Comput. Biol.* **1**, e34 (2005).
7. P. Murray-Rust, H. S. Rzepa, *J. Digit. Inform.* **5**, issue 1 (2004).
8. D. Nicholas, P. Huntington, P. Williams, T. Dobrowolski, *J. Doc.* **60**, 24 (2004).
9. D. Nicholas, P. Huntington, H. R. Jamali, T. Dobrowolski, *Inf. Process. Manage.* **43**, 1085 (2007).
10. C. Tenopir, D. W. King, *D-Lib Mag.* **14**, issue 11/12 (2008).
11. B. Schatz *et al.*, *Computer* **32**, 51 (1999).
12. A. P. Bishop, *Inf. Process. Manage.* **35**, 255 (1999).
13. C. L. Palmer, *Work at the Boundaries of Science: Information and the Interdisciplinary Research Process* (Kluwer, Dordrecht, Netherlands, 2001).
14. C. L. Palmer, M. H. Cragin, T. P. Hogan, *Inf. Process. Manage.* **43**, 808 (2007).
15. O. Bodenreider, R. Stevens, *Brief. Bioinform.* **7**, 256 (2006).
16. L. Strömback, D. Hall, P. Lambrix, *Proteomics* **7**, 857 (2007).
17. M. Ashburner *et al.*, *Nat. Genet.* **25**, 25 (2000).
18. B. Smith *et al.*, *Nat. Biotechnol.* **25**, 1251 (2007).
19. Gene Ontology, www.geneontology.org/.
20. The examples are from “An Introduction to the Gene Ontology” www.geneontology.org/GO.doc.shtml.
21. World Wide Web Consortium, www.w3.org/XML/.
22. J. H. Coombs, A. H. Renear, S. J. DeRose, *Commun. ACM* **30**, 933 (1987).
23. A. Renear, D. Dubin, C. M. Sperberg-McQueen, C. Huitfeldt, in *Proceedings of the 2002 ACM Symposium on Document Engineering*, R. Furuta, J. I. Maletic, E. Muson, Eds. (Association for Computing Machinery Press, New York, 2002), pp. 119–126.
24. World Wide Web Consortium, www.w3.org/2001/sw/.
25. J. A. Blake, C. J. Bult, *J. Biomed. Inform.* **39**, 314 (2006).
26. J. Blake, *Nat. Biotechnol.* **22**, 773 (2004).
27. M. R. Seringhaus, M. B. Gerstein, *BMC Bioinform.* **8**, 17 (2007).
28. D. Sholton, *Learn. Publ.* **22**, 85 (2009).
29. M. Krallinger, A. Valencia, *Genome Biol.* **6**, 224 (2005).
30. J.-j. Kim, D. Rebolz-Schuhmann, *Brief. Bioinform.* **9**, 452 (2008).
31. H. M. Müller, E. E. Kenny, P. W. Sternberg, *PLoS Biol.* **2**, e309 (2004).
32. R. Hoffmann, A. Valencia, *Nat. Genet.* **36**, 664 (2004).
33. I. Foster, *Science* **308**, 814 (2005).
34. P. Boyce, D. W. King, C. Montgomery, C. Tenopir, *Ser. Libr.* **46**, 121 (2004).
35. D. W. King, C. Tenopir, C. Montgomery, S. E. Aerni, *D-Lib Mag.* **9**, issue 10 (2003).
36. We thank G. Bilder (Crossref), M. Sperberg-McQueen (Black Mesa Technologies), M. Smith (MIT Libraries), and W. J. MacMullen and L. C. Smith (University of Illinois) for helpful comments on an earlier version of this paper and L. Tefreau (University of Illinois) for assistance with the examples and illustrations. Earlier versions were presented at the 2006 STM Innovations Seminar (London), the 2007 Annual Meeting of the American Chemical Society (Chicago), the 2007 STM Spring Conference (Cambridge, MA), and the 2007 annual meeting of the American Library Association (Washington, DC). This research was supported in part by NSF grant 0222848.

10.1126/science.1157784