Knowledge Encapsulation Framework for Collaborative Social Modeling

Andrew J. Cowell, Michelle L. Gregory, Eric J. Marshall, Liam McGrath

Pacific Northwest National Laboratory
902 Battelle Boulevard, P.O. Box 999, Richland, WA 99352, USA
{andrew.cowell | michelle.gregory | eric.marshall | liam.mcgrath}@pnl.gov

Abstract

This paper describes the Knowledge Encapsulation Framework (KEF), a suite of tools to enable knowledge inputs (relevant, domain-specific facts) to modeling and simulation projects, as well as other domains that require effective collaborative workspaces for knowledge-based task. This framework can be used to capture evidence (e.g., trusted material such as journal articles and government reports), discover new evidence (covering both trusted and social media), enable discussions surrounding domain-specific topics and provide automatically generated semantic annotations for improved corpus investigation. The current KEF implementation is presented within a wiki environment, providing a simple but powerful collaborative space for team members to review, annotate, discuss and align evidence with their modeling frameworks. The novelty in this approach lies in the combination of automatically tagged and user-vetted resources, which increases user trust in the environment, leading to ease of adoption for the collaborative environment.

Introduction

Researchers across all domains in academia, industry and government, have the onerous task of keeping up with literature in the fields of study and related fields. The use of the Internet has made long distance collaborations possible and thus has increased productivity of researchers in general. In addition, the Internet makes it easier for academic journals, conferences, workshops, and individual researchers to put the fruits of their labor in front of a larger audience. The Internet has also made it easier than ever to perform searches and find relevant information.

However, the use of the Internet as a research tool has its limitations due to the quantities of data available and often questionable quality (not to mention the multitude of file formats and standards). In the sea of Adobe PDF and Microsoft Word files that take up space on their (electronic) desktop, researchers are finding it more difficult to identify relevance and significance of individual articles in the mass of similarly titled material. Once material is found, the benefits of electronic media end there: researchers are still more comfortable printing out relevant documents and making notes in margins. Additionally, researchers will send links for electronic documents to their collaborators and each will individually print and make margin annotations. It is not uncommon for intelligence analysts (a specific type of knowledge worker that the authors have experience with), to spend 80% of their time collecting material for their tasking, leaving only 20% of time for the analysis. In the research described herein, we aim to address both the quantity of data problem as well as making use of electronic media to increase collaboration and productivity. We do this through a collaborative wiki environment designed to find and filter input data, allow for user input and annotations, and provide a workspace for team members. This system is also designed to link data from sources directly to a research area for maximum productivity and pedigree. In this manner, we're hoping to utilize a 'wisdom of the crowds' approach to even out collection and analysis time and effort to a more reasonable ratio.

In this paper, we describe the Knowledge Encapsulation Framework (KEF). After a discussion of prior art, we describe the system concept followed by current implementation details and finally a use-case from the climate change domain.

Prior Art

The fundamental concept for KEF has been investigated across a number of disciplines for a number of years. Experts systems (Ignizio 1991; Jackson 1998) research have tried to capture the tacit knowledge residing within a specific domain (usually through the elicitation of that knowledge from subject matter experts) so that this information can be shared and transferred to other members. Our work does not attempt to codify or understand the knowledge that an SME brings to a problem. The KEF environment simply provides a collaborative environment where such individuals can collectively discuss and discover new facts within a dynamic stream of incoming information. In addition, a common interface to an expert system is to consider it to *be* an expert that can answer questions either through a traditional text-based interface or a more anthropomorphic representation that may appear to have human form and that can listen and talk to the user

(Cowell and Stanney 2004). KEF, on the other hand, is simply an environment that allows for the discussion and evolution of new knowledge and ideas. There is also often a significant amount of effort placed in engineering the knowledge structure in expert systems so that reasoning can occur to handle unforeseen situations. While KEF does attempt to annotate semantic relationships identified within the data sources, these are not hard-coded ontologies – rather, we build up a categorization scheme based on the content identified. Finally, typical expert systems focus on a very narrowly defined domain, such as Mycin and CADUCEUS (both medical diagnosis systems), NeteXPERT (network operations automation system), KnowledgeBench (new product development applications) and Dipmeter Advisor (oil exploration system). The concepts set out for KEF can be generalized for any domain.

Collaborative problem solving environments (CPSE) are perhaps a better analogy for the concept KEF is attempting to convey. The Pacific Northwest National Laboratory has a long history of building CPSE's for Department of Energy (DOE) scientists (PNNL 2002), such as the DOE2000 Electronic Notebook Project (Myers 1996). Watson (2001) reviewed a number of organizations pursuing CPSE's including other DOE sites (e.g., Common Component Architecture, Collaboratory Interoperability Framework, Corridor One Project) as well as the Department of Defense (e.g., Gateway), NASA (e.g., Intelligent Synthesis Environment (ISE), Collaborative Engineering Environment (CEE) and Science Desk) and numerous university efforts (Rutgers University's Distributed System for Collaborative Information Processing and Learning, University of Michigan's Space Physics and Aeronomy Research Collaboratory and Stanford's Interactive Workspaces). Shaffer (2006), in his position statement on CPSE's defined them as a "system that provides an integrated set of high level facilities to support groups engaged in solving problems from a proscribed domain". These facilities are most often directly related to the domain, e.g., facilities to enable 3D molecular visualization for biologists. KEF includes a number of components but the focus has always been on the general case – i.e., development of capabilities that apply across a number of domains. Within CPSE's, there is also significant amount of effort placed in encouraging synchronous interaction, a facility provided by KEF through an integrated textual chat component but secondary to the asynchronous wiki implementation. Perhaps the most striking difference between traditional CPSE's and our implementation of KEF is the scale of effort. Many of the CPSEs mentioned above were created over a number of years at the cost of millions of dollars, and have an excessive learning curve and setup time. KEF, while leveraging the experiences of these previous systems, is built using open-source software (e.g., the same wiki framework used in Wikipedia[1]) and is configurable within a few hours.

Perhaps the most similar technology currently available to KEF are the 'web 2.0' information stores available on the Internet. Examples include encyclopedic resources such as Wikipedia and Knol[2] that rely on 'wisdom of the crowds' to build and maintain a knowledge base of information. Such resources rarely utilize automated processes to extract semantic relations and add these as additional metadata that can aid in the discovery process[3]. Like KEF, some of these systems use tags to provide an informal tagging mechanism but the domain scale are typically very wide (in the case of Wikipedia, the goal is to provide an encyclopedia's worth of knowledge). Project Halo (Friedland et al. 2004) is specific instance of an information store that aims to develop an application capable of answering novel questions and solving advanced problems in a broad range of scientific disciplines (e.g., biology, physics, and chemistry). The mechanism for inserting knowledge into the data store (i.e., using graduate students with domain knowledge) requires significant effort. The KEF approach is to share the load between automated information extraction tools and domain experts (as see in Figure 1). While we acknowledge the limitations of automated information extraction technologies, we believe an approach that leverages automated means while encouraging users to make corrections and provide their own annotations may provide significantly rich metadata.

System Concept & Design

The fundamental concept behind KEF is of an environment that can act as an assistant to a research team. By providing some documents (e.g., research articles from a domain of focus) as an indication of interest, elements of the KEF environment can automatically identify new and potentially related material, inserting this back into the environment for review. KEF can be configured to harvest information from individual sites, use search engines as proxies, or collect material from social media sites such as blogs, wikis, and forums etc. Harvesting strategies include:

- simple metadata extraction (e.g., author and co-author, material source (e.g., journal name), citations within original documents, etc)
- topic identification (e.g., climate-change, food supply, access to education, etc)
- sentiment analysis (e.g., the fact that the statements related to climate-change are positive or negative)

[1] http://www.wikipedia.org
[2] http://knol.google.com
[3] Although a new effort entitled DBpedia (http://dbpedia.org) is a community effort to extract structured information from Wikipedia.

- rhetorical analysis (e.g., identification of issues being relayed from a protagonist to a target audience with a specific intent to cause an effect).

Initial results may lack close relevance due to the general criteria for search. Users can vet the material collected, either by single items or by groups (e.g., everything from a particular author or journal). This procedure serves as input to the harvesting strategy until a tightly defined harvesting strategy matches exactly with what the research team needs. Eventually, the research team can expect to receive a steady stream of relevant traditional material and social media.

Figure 1. The KEF Shared Effort Concept

As the data repository is populated with relevant material, users can interact with the data on a variety of levels depending on their goals. All data in the repository is automatically tagged with basic document metadata (source, author, date, etc.), as well as with semantic information extracted from the text during the ingestion routine. Using information extraction tools, all entities (people, locations, events, etc.) in the text are marked and user-identified key terms are automatically tagged (e.g., climate terms in the case of a climate modeling scenario). These tags provide a means of search and organization that provide for ease of recall. Importantly, users can correct existing annotations, or create their own to match their individual needs. Users can replace manual margin mark-up with notes or annotations that can be searched on later or used by other collaborators. Finally, each document has a talk page where users can discuss (asynchronously) the document. A synchronous 'chat' component is also available.

The Process

From a users perspective, the KEF process is illustrated in Figure 2. Knowledge elicitation experts meet with modelers and subject-matter experts[4] to get an understanding of their problem. For example, in the case of a modeling group trying to understand the effects of climate change on the Indian sub-continent, this may lead to the creation of a context map showing all the elements of climate change that may apply (e.g., access to education, clean water, etc) and a selection of documents currently used to create and parameterize their models.

[4] Depending on the domain, these may be the same person.

Documents collected in this first phase are used as part of the discovery phase. The documents are 'virtually' dissected by a number of KEF components (i.e., automated software tools) in order to understand their constituents and relevance. Based on these elements, new material (e.g., documents, websites, blogs, forums, news articles, etc) are discovered and pushed through an extraction pipeline prior to being ingested into the knowledge base. This process is cyclic, altered by the feedback provided by the user during the vetting/review phase.

As material is introduced to the knowledge base, it can be reviewed by the users through the KEF wiki. The wiki provides a simple but powerful collaborative environment for the vetting, evaluation and alignment of evidence to models. Each of these phases is described in more detail in the following sections.

Phase 1: Knowledge Elicitation

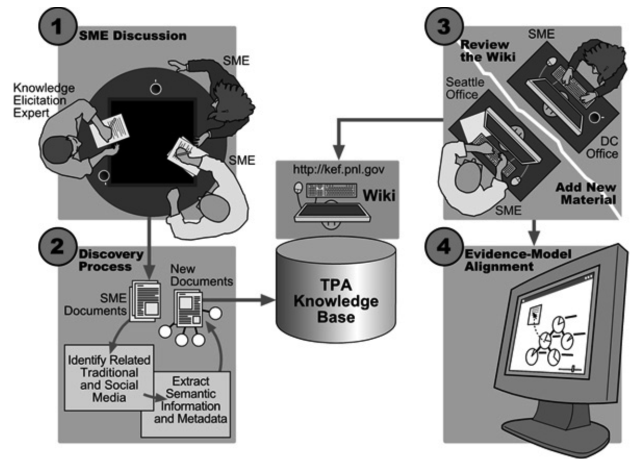


Figure 2. The KEF Process

In order to initiate the automatic harvest-user vetting cycle, we need an understanding of the intended users' domains. The extraction of expert knowledge so that it can be actively utilized by non-experts (in this case, an automated system) has been the focus of a number of high profile research projects. Perhaps the most widely known of these efforts was the DARPA Rapid Knowledge Formation (RKF) Program that attempted to enable distributed teams of users to enter and modify knowledge directly without the need for specialized training in knowledge representation, acquisition, or manipulation. A more recent ancestor of the RKF program was mentioned earlier - Project Halo (Friedland et al. 2004), an ambitious endeavor that aims to develop an application capable of answering novel questions and solving advanced problems in a broad range of scientific disciplines. Some elements of Project Halo are already being reused within the KEF repository (e.g., the mediawiki engine with semantic extensions). While many types of knowledge elicitation techniques exist (Burge

2005), in KEF we rely on structured interviewing (Hudlicka 1997) with case study discussions directly related to the selected domains (Geiwitz et al. 1990). We have also investigated the use of concept mapping (Thordsen 1991; Gowin and Novak 1984) as part of the structured interviews in order to gather a graphical representation of the scenario.

In addition to structured interviews, users are asked to provide a list of trusted sources (e.g., specific journal articles, government reports, etc) that they rely on for building and/or parameterizing their models. In some cases this may result in large amounts of printed matter that that needs to be ingested into the framework using an optical character recognition (OCR) workflow. These trusted sources initiate the automatic discovery process (Figure 3).

Phase 2: The Discovery Process

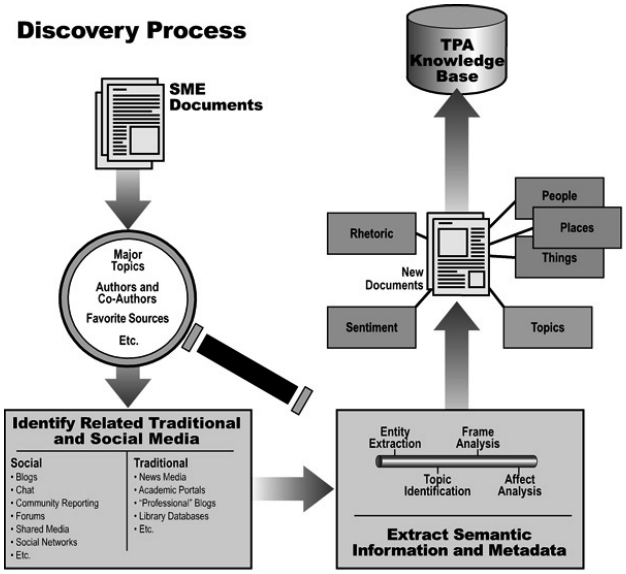


Figure 3. The KEF Discovery Process

In order to isolate relevant parts of information from our user-supplied sources, we have employed a number of automated information extraction tools. These tools provide a search template that directs the discovery process (for example, extracted topics from the user-supplied sources that are then used in conjunction with document metadata to return new, potentially relevant documents). In addition, the same components are used to annotate the new material to provide multiple levels of summary and visualization through the vetting process.

The new documents are placed directly into the KEF knowledge base, but tagged to show their unvetted status. They are user-accessible through the KEF wiki, with each document representing a page in the wiki.

Phase 3 & 4: Reviewing/Aligning Material

While wiki's are generally repositories of information, albeit dynamically generated through the interaction of multiple users, one key element of this framework is the interactive manipulation of knowledge within the environment. As the discovery process identifies new material, the environment leads the user through a four-stage analysis; review, relevance, evaluation and task alignment. These stages can be viewed in Figure 4.

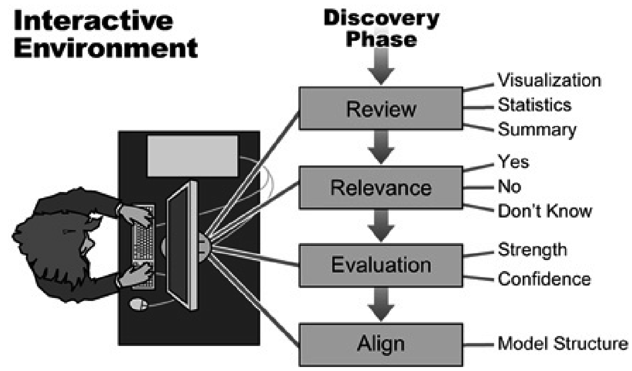


Figure 4. The KEF Review Process

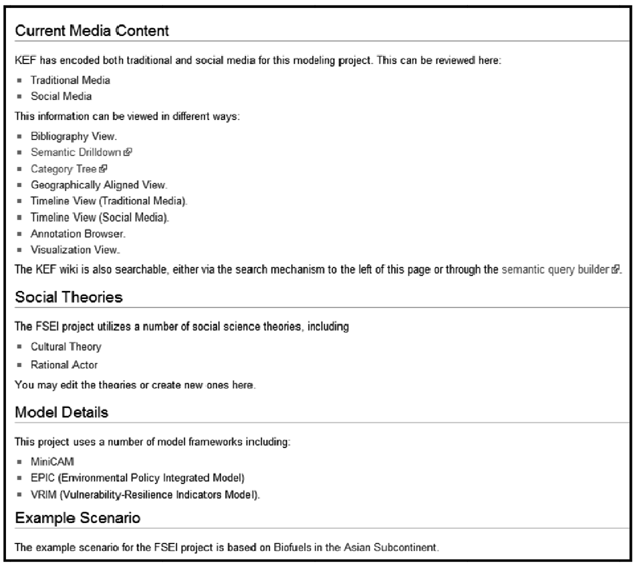
1. Review We use multiple automated methods to describe the new content including statistics (e.g., 40% of new material came from blogs), simple easy-to-understand graphic visualizations (e.g., pie and bar chart representations, as well as more complex clustering representations) and material summaries. The aim is to provide enough context so that the user can decide on material relevance without having to read the entire document (although a link to the complete text is provided, should the user require it).

2. Relevance For each piece of evidence, the user is required to make a judgment regarding its relevance to their current domain. We provide three alternatives: 'Yes' (the document is directly relevant to my problem), 'No' (the document is irrelevant to my problem) and 'Don't Know'. A 'Don't Know' response might be due to the individual not having the expertise to make that decision (e.g., it is outside their domain) or simply due to them not having time to make the decision. Any response removes the item from their work list although it remains within the lists for their colleagues in order to capture multiple opinions. Material that receives a large number of irrelevant votes is moved to an archival namespace and is no longer included in system statistics (although it can still be retrieved by users if they wish).

3. Evaluation A response of 'Yes' leads the user through two more dialogs. In the evaluation stage, the user rates the document importance (valence/strength) with respect to their task and the document credibility (i.e., a subjective measure of trust in the document contents). We aim to use these ratings in a recommendation context ('users similar to yourself also liked…') and for training the discovery system. For certain modeling systems (e.g., Bayesian analysis), these ratings may also be used to define node strength.

4. Task Alignment After evaluating the document, the final stage allows the user to align the data to any structure relevant to their task. In the case of a model, the user may align the material directly to a specific model input. For other applications, this may be a simple sorting mechanism for further interpretation. Concept maps generated in the first phase might provide a sufficient mechanism for alignment.

Technical Implementation


Current Media Content

KEF has encoded both traditional and social media for this modeling project. This can be reviewed here:

- Traditional Media
- Social Media

This information can be viewed in different ways:

- Bibliography View.
- Semantic Drilldown
- Category Tree
- Geographically Aligned View.
- Timeline View (Traditional Media).
- Timeline View (Social Media).
- Annotation Browser.
- Visualization View.

The KEF wiki is also searchable, either via the search mechanism to the left of this page or through the semantic query builder.

Social Theories

The FSEI project utilizes a number of social science theories, including

- Cultural Theory
- Rational Actor

You may edit the theories or create new ones here.

Model Details

This project uses a number of model frameworks including:

- MiniCAM
- EPIC (Environmental Policy Integrated Model)
- VRIM (Vulnerability-Resilience Indicators Model).

Example Scenario

The example scenario for the FSEI project is based on Biofuels in the Asian Subcontinent.

Figure 5. Access Points to Review Material

The current implementation is based on a Mediawiki[5] installation. Mediawiki is arguably the most well known of the current generation of wiki engines and is the engine used in both Wikipedia[6] and Intelipedia[7]. Due to its popularity, a number of third-party extensions have been developed, and, where possible, we have leveraged these to provide a more effective environment. For example, instead of implementing our own semantic layer where relationships can be defined within document elements (e.g., Paul is_the_author_of Document-X), we employed the Semantic Mediawiki (SMW[8]) extension.

A number of access views are used to enable users to interact with the information in the repository using their preferred mechanism. For example, Figure 5 shows part of a project page with a number of access points to access information (e.g., through a timeline (Figure 6), geographically aligned (Figure 7) or in a table (Figure 8)).

The automated information extraction pipeline consists of three main components: a named entity recognizer, a summarizer, and a geo-reference component. The named entity recognizer (NER) is based on the Connexor[9] software that identifies proper names, selects common nouns and maps them to semantic tags via their identified categories. Connexor Metadata[10] is used to identify the names, their categories and their lemmatized form. The summarizer component, based on the publicly available MEAD summarizer[11], creates a summary of one or more pages of a source document by identifying the most significant sentences. It is used in the pipeline to create 'summary' properties for single-document pages. Finally, the geo-reference component provides a map view of source documents based on NER locations. Locations are

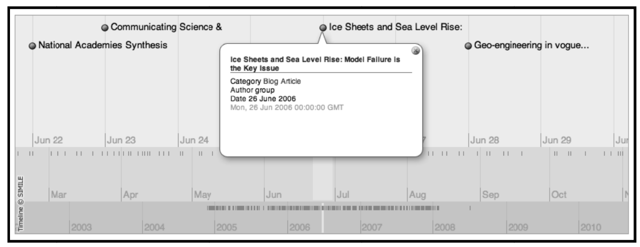


Figure 6. The Timeline View

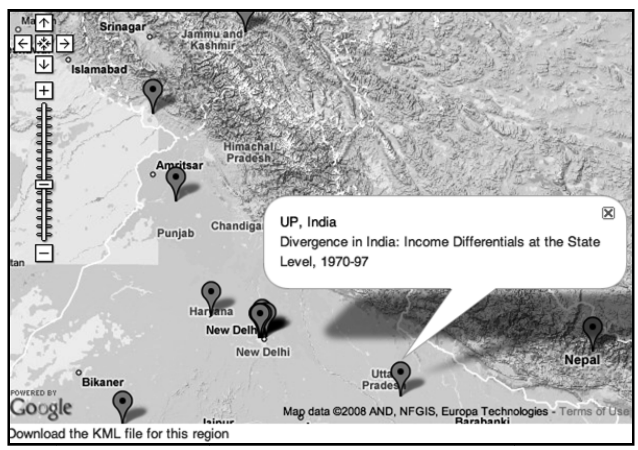


Figure 7. The Geographic View

[5] http://www.mediawiki.org
[6] http://www.wikipedia.org
[7] http://en.wikipedia.org/wiki/Intellipedia
[8] http://www.semantic-mediawiki.org

[9] http://www.connexor.com/
[10] http://www.connexor.eu/technology/machinese/
[11] http://www.summarization.com/mead/

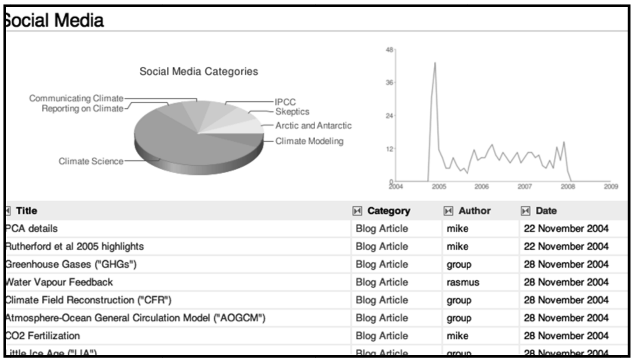


Figure 8. The Tabular View

checked against the Google Geocoding Service[12] to create a set of coordinates associated with an article. For pages with many location names, these coordinates can be restricted to those that occur most frequently, or those in the summary or title of the page to highlight the most relevant locations. After a set of pages has been processed, a new page is created with the map view using the Google Maps MediaWiki extension. This component creates a 'Geographically Aligned View' page in the wiki (as shown above in Figure 7).

Finally, to provide a compelling user experience, we implemented a custom version of the Project Halo[13] extension. This allows the user to adjust the semantic markup made by the automated discovery mechanism and also make their own annotations.

Figure 9 shows how a document may appear within the repository. The citation and an offline version (in this example, an Adobe PDF file) are shown at the top of the page, with an 'infobox' showing the journal the paper came from and other metadata (e.g., author, date, etc). A link is maintained to the original source URL. The text within the page can not be altered although users can add their annotations (Figure 10) and review the automated and manual annotations (Figure 11).

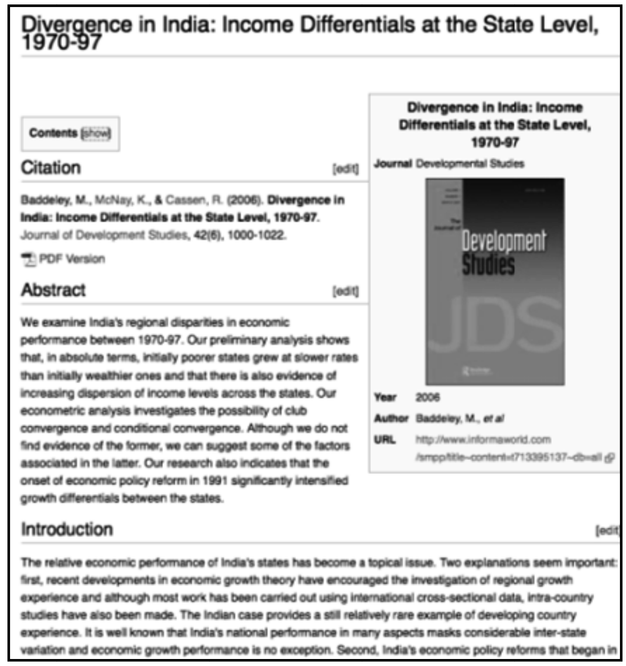


Figure 9. A Journal Article

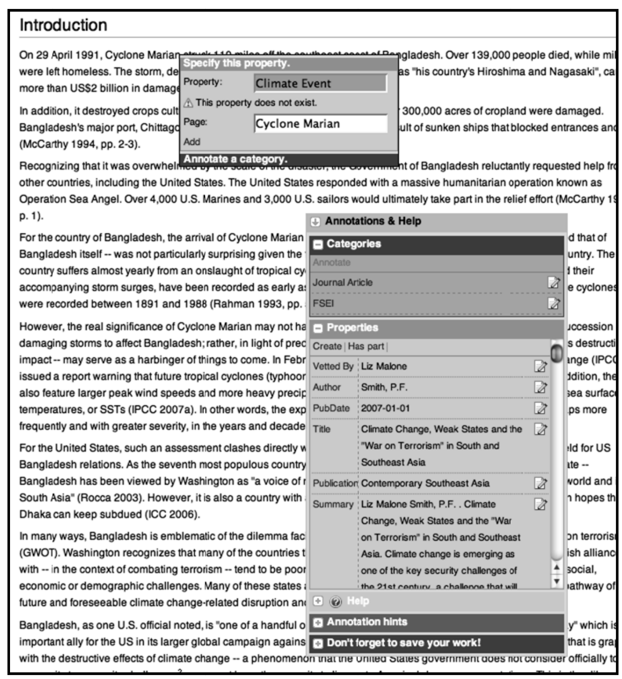
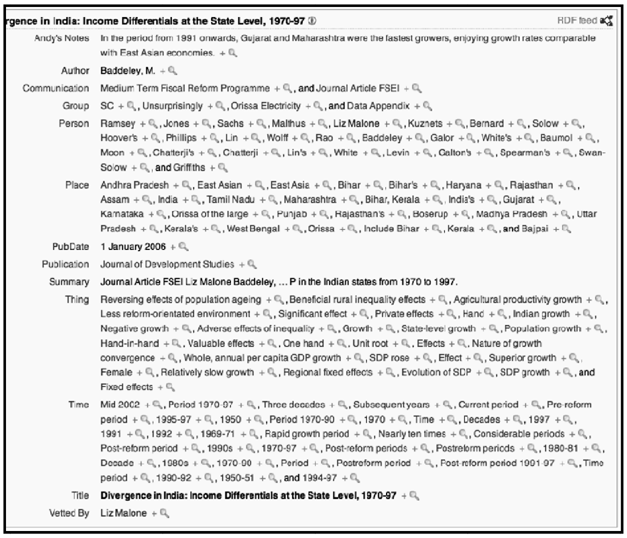


Figure 10. Annotating Text in the Wiki

[12] http://code.google.com/apis/maps/documentation/services.html
[13] http://semanticweb.org/wiki/Project_Halo



rgence in India: Income Differentials at the State Level, 1970-97 RDF feed

Andy's Notes	In the period from 1991 onwards, Gujarat and Maharashtra were the fastest growers, enjoying growth rates comparable with East Asian economies.
Author	Baddeley, M.
Communication	Medium Term Fiscal Reform Programme, and Journal Article FSEI
Group	SC, Unsurprisingly, Orissa Electricity, and Data Appendix
Person	Ramsey, Jones, Sachs, Malthus, Liz Malone, Kuznets, Bernard, Solow, Hoover's, Phillips, Lin, Wolff, Rao, Baddeley, Galor, White's, Baumol, Moon, Chatterji's, Chatterji, Lin's, White, Levin, Galton's, Spearman's, Swan-Solow, and Griffiths
Place	Andhra Pradesh, East Asian, East Asia, Bihar, Bihar's, Haryana, Rajasthan, Assam, India, Tamil Nadu, Maharashtra, Bihar, Kerala, India's, Gujarat, Karnataka, Orissa of the large, Punjab, Rajasthan's, Boserup, Madhya Pradesh, Uttar Pradesh, Kerala's, West Bengal, Orissa, Include Bihar, Kerala, and Bajpai
PubDate	1 January 2006
Publication	Journal of Development Studies
Summary	Journal Article FSEI Liz Malone Baddeley, ... P in the Indian states from 1970 to 1997.
Thing	Reversing effects of population ageing, Beneficial rural inequality effects, Agricultural productivity growth, Less reform-orientated environment, Significant effect, Private effects, Hand, Indian growth, Negative growth, Adverse effects of inequality, Growth, State-level growth, Population growth, Hand-in-hand, Valuable effects, One hand, Unit root, Effects, Nature of growth convergence, Whole, annual per capita GDP growth, SDP rose, Effect, Superior growth, Female, Relatively slow growth, Regional fixed effects, Evolution of SDP, SDP growth, and Fixed effects
Time	Mid 2002, Period 1970-97, Three decades, Subsequent years, Current period, Pre-reform period, 1995-97, 1950, Period 1970-90, 1970, Time, Decades, 1997, 1991, 1992, 1969-71, Rapid growth period, Nearly ten times, Considerable periods, Post-reform period, 1990s, 1970-97, Post-reform periods, Postreform periods, 1980-81, Decade, 1980s, 1970-90, Period, Postreform period, Post-reform period 1991-97, Time period, 1990-92, 1950-51, and 1994-97
Title	Divergence in India: Income Differentials at the State Level, 1970-97
Vetted By	Liz Malone

Figure 11. The Annotations for a Journal Article

Use Case

To illustrate the process from a user perspective, we have built a use case describing the environment applied to a Climate Change & Security scenario. Not all the functionality described in the use case has been implemented at this time; however, it serves as an effective vision for what KEF could potentially provide knowledge-workers in the near future.

Liz, the lead for the 'Vulnerability of Food Security and Energy Infrastructures to Climate Change and Terrorism' project, opens her browser window and selects a bookmark for the Knowledge Encapsulation Framework Wiki. Liz runs her cursor down the navigation menu and selects her project.

On clicking the link, Liz is presented with her project's homepage on the wiki. In the top right-hand corner some usage statistics are shown. She notices that her graduate student logged on yesterday at 4.23pm. She can see what he did and tasks he has prepared for her. The main section of the page holds details on new material that has arrived since she last logged on to her project wiki. Today, it appears a new blog has been detected that might be of interest. There are also 32 blog articles and 10 news articles related to concepts Liz has previously defined as important. Below this space is an area that describes the current state of her modeling effort. This includes a concept map of the main components of her model, and important terms (either selected by Liz or suggested by the system and accepted by Liz). She feels the model is specifically weak on information related to food supply in India.

She moves her cursor to the 'new items' space and clicks 'explore'. This takes her to a new page that lists the new items in a tabular form. She requests to see the new material graphically. The wiki presents the new material geographically aligned with the locations mentioned in the articles, and also as a tag cloud with most common terms appearing larger than those of less importance. A timeline view also presents all 42 articles in publication order. She notices both 'Food' and 'India/Indian' appear prominently in the cloud. She restricts the geographical visualization to reports mentioning India, and then clicks 'Food' in the tag cloud. The table refreshes to only show 5 articles.

Liz clicks on the first and is provided with a summary of the article. It appears to be only tangentially related, so she rejects it and moves on to the next article. This one appears to be much more interesting. After reading the summary, Liz clicks the link to see the full article, as presented within the wiki. The page opens and the 'infobox' in the top right hand corner displays summary details. Terms extracted from the document are shown at the bottom of the article (such as PLACES: India, Pakistan and PEOPLE: John Smith). This provides another type of summary for Liz to get an idea of how useful the article might be without reading line by line. She is interested, however, and chooses to review the entire article. She even fixes a couple of errors made by the automated annotation system.

There is one specific paragraph and table that refers directly to her problem domain. It describes crop estimates for the different regions in India. She annotates this by attaching the FOOD property tag to the paragraph and table. She notes, however, that it does not match estimates made by another research article. She annotates the article and the comparative article to show the differences. She may send both articles to Cesar (her colleague) or another project team member to help determine which estimates are more "trusted." Or she may link the annotations so that the model may use either or a range derived from both.

After reviewing the rest of the articles and the new blog that was identified by KEF (narrowing the search parameters for the blog so that only articles relating to India, Pakistan, and Bangladesh are harvested), Liz moves through into her vetted materials space. This page describes those documents, articles and miscellaneous notes that she has collected that are important to her model. In this space, Liz can write notes, make annotations and start to make linkages between the material and specific elements of her model. She can also start to make assessments of the strength, relevance and credibility of the material (e.g., a particular article might be very relevant to a particular model node, and be strongly supporting on side of the argument but come from a source that isn't very credible).

She notices that Cesar had added some new material to the Governance node. She is able to review the particular piece of the government report he annotated. She makes an assessment of the strength and credibility of the document that is significantly lower than those Cesar placed, as she knows the author and remembers he has a tendency to over-estimate. She clicks on his name (which creates a new wiki page specifically for this author) and writes a quick note to explain why. She then reviews her "notes and questions" space, adds several comments and a question to it.

Time is getting on and Liz needs to get on the road to avoid the afternoon traffic. She logs off the wiki, retrieves a printed article from the printer, and makes for the door....

Conclusion

We have presented a collaborative workspace for researchers to gather, annotate and store relevant information. The combination of automatically harvested material with user vetting helps the researcher effectively handle the potentially large quantities of data available, while providing a measure of quality control.

The use of a semantic wiki allows multiple people to add, vet and discuss source material, enabling effective collaboration. The co-location of source material, user annotation and discussions provides for effective collection of provenance.

Acknowledgements

The authors would like to acknowledge the support and feedback of the Technosocial Predictive Analytics Initiative at the Pacific Northwest National Laboratory as well as the reviewers and editors that have helped refine this paper and the associated research.

References

Besancon, R., M. Rajman, et al. 1999. Textual similarities based on a distributional approach. Database and Expert Systems Applications, Florence, Italy, IEEE.

Geiwitz, J., Kornell, J., McCloskey, B. 1990. An Expert System for the Selection of Knowledge Acquisition Techniques. Technical Report 785-2, Contract No. DAAB07-89-C-A044. California, Anacapa Sciences.

Gowin, R., Novak, J.D. 1984. *Learning how to learn*. NY: Cambridge University Press.

Hammouda, K. M., Kamel, M.S. 2004. Document Similarity Using a Phrase Indexing Graph. *Knowledge and Information Systems* 6(6): 710-727.

Hudlicka, E. 1997. Summary of Knowledge Elicitation Techniques for Requirements Analysis, Course Material for Human Computer Interaction, Worcester Polytechnic Institute.

Pearl, J. 1988. Probabilistic Reasoning in Intelligent Systems, San Fransisco, CA: Morgan Kaufmann.

Purvis, N. and J. Busby. 2004. The Security Implications of Climate Change for the United Nations System. ECSP Report, Issue 10.

Schafer, G. 1976. *A Mathematical Theory of Evidence*. Princeton, NJ: Princeton University Press.

Schum, D.A. 2001. The Evidential Foundations of Probabilistic Reasoning. Evanston, IL: Northwestern University Press.

Select Committee on Intelligence. 2004. Report on the U.S. Intelligence Community's Prewar Intelligence Assessments on Iraq. Washington, D.C. United States Congress.

Thordsen, M. 1991. A Comparison of Two Tools for Cognitive Task Analysis: Concept Mapping and the Critical Decision Method. In Proceedings of the Human Factors Society 35th Annual Meeting, 283-285.

Ignizio, J.P. 1991. *Introduction to Expert Systems: The Development and Implementation of Rule-Based Expert Systems*, McGraw-Hill.

Jackson, P. 1998. *Introduction to Expert Systems (3rd Edition)*. Addison Wesley.

PNNL. 2002. Collaborative Problem Solving Environments. Available from http://www.pnl.gov/cpse/whatwedo.stm. Accessed September 22nd, 2008.

Shaffer, C.A. 2006. Collaborative Problem Solving Environments. Available from http://people.cs.vt.edu/~shaffer/Papers/DICPMShaffer.html. Accessed October 2nd, 2008.

Myers, J.D. et al. 1996. Electronic Laboratory Notebooks for Collaborative Research. In Proceedings of WETICE '96. Stanford University, June 19-21, 1996. IEEE Computer Society, 47-51.

Watson, V.R. 2001. Supporting Scientific Analysis within Collaborative Problem Solving Environments. In HICSS-34 Minitrack on Collaborative Problem Solving Environments, Maui, Hawaii, January 3-6, 2001.

Cowell, A.J., Stanney, K. 2004. Manipulation of Non Verbal Interaction Style and Demographic Embodiment to Increase Anthropomorphic Computer Character Credibility. International Journal of Human-Computer Studies, 62(2) : 281-306.