Security Informatics
a SpringerOpen Journal

RESEARCH Open Access

Knowledge encapsulation framework for technosocial predictive modeling

Michael C Madison[1*], Andrew J Cowell[1], R Scott Butner[1], Keith Fligg[1], Andrew W Piatt[1], Liam R McGrath[1] and Peter C Ellis[2]

Abstract

Analysts who use predictive analytics methods need actionable evidence to support their models and simulations. Commonly, this evidence is distilled from large data sets with significant amount of culling and searching through a variety of sources including traditional and social media. The time/cost effectiveness and quality of the evidence marshaling process can be greatly enhanced by combining component technologies that support directed content harvesting, automated semantic annotation, and content analysis within a collaborative environment, with a functional interface to models and simulations. Existing evidence extraction tools provide some, but not all, the critical components that would empower such an integrated knowledge management environment. This paper describes a novel evidence marshaling solution that significantly advances the state of the art. Its embodiment, the Knowledge Encapsulation Framework (KEF), offers a suite of semi-automated and configurable content harvesting, vetting, annotation and analysis capabilities within a wiki-enabled and user-friendly visual interface that supports collaborative work across distributed teams of analysts. After a summarization of related work, our motivation, and the technical implementation of KEF, we will explore the model for using KEF and results of our research.

Keywords: Semantic web, Technosocial predictive analytics, Predictive analytics, Knowledge management, Knowledge encapsulation framework, Semantic MediaWiki, Web-based interaction, Collaborative computing environments, Data mining, Web harvesting, Natural language processing

Introduction

Information analysts and researchers across many domains in academia, industry, and government have the onerous task of culling and searching through large data sets of traditional and social media to support their research in their domain. While the internet has simplified distance collaboration and increased many facets of an individual's or team's productivity [1], it has also significantly increased the number of possible traditional (e.g., journal articles, conference papers, technical reports, etc.) and social media (e.g., blogs, Twitter, etc.) sources that the analyst must locate, fact check, and leverage in a meaningful way [2]. The *Washington Post* helps to illustrate the quantity of data that can be accumulated rapidly when social media is combined with traditional media surrounding a topic with a recent blog post focusing on Twitter volume during the 2012 Republican Presidential primary. Well over 200,000 tweets were made about the front-running 2012 Republican candidates in *a single day* [3]. This is, of course, insignificant to the amount of social media data churned out daily by Twitter alone (approximately 140,000,000 tweets per day as of this writing [4]), in addition to Facebook, LinkedIn, Google+, and the other prominent social networking sites.

The analyst, who's research is enabled by this mountain of data, is now responsible for combining the various sources, fact checking each record, marshaling the evidence, and aligning it with models for predicting future events. This analyst's job can be made simpler through the use of state-of-the art data mining and harvesting applications, which can automatically locate and combine disparate data repositories into a single, much larger repository. The analyst can then go to a single location to search for relevant evidence instead of searching multiple locations. Once harvested, these data can

* Correspondence: michael.madison@pnnl.gov
[1]Pacific Northwest National Laboratory, 902 Battelle Boulevard, 999, MSIN K7-28 Richland, WA 99352, USA
Full list of author information is available at the end of the article

be fed through other analytical applications to find relevant named entity, location, or event mentions that would be of interest to the analyst or the models. Finally, the analyst can use existing collaborative tools to interact with peers who might be doing similar research. Unfortunately, these solutions have not yet been integrated into a single tool, leaving much of the burden on the analyst for moving data between applications and noticing relevant information once it's been collected. For example, environments such as IBM SPSS and SAS Analytics, which are the instruments of choice for predictive analysis, provide tools for data collection through surveys, data mining, and presentation but do not offer a collaborative framework with capabilities for harvesting content from the internet and automated semantic annotation.

Why does a predictive analyst need such powerful features combined in a single tool? Consider the diversity of the research that a predictive analyst might face in today's world:

- What does the use of social media tools such as Facebook and Twitter in the recent "Arab Spring" uprisings tell us about the regimes in the region that are most vulnerable to similar rebellions? How might cultural differences affect the translation of these phenomena to other parts of the world?
- Assuming that the high incidence of 100-degree days in much of the southern United States during the summer of 2011 is a long-term trend, what are the likely implications for U.S. power grid operations? Will any of the anticipated changes in electrical load create new vulnerabilities in the grid? Where are these vulnerabilities likely to be concentrated? How might they be mitigated?
- How would one recognize the "early warning signs" of an emerging terrorist network that has the goal of building a nuclear weapon? How could these warning signs be differentiated from activities resulting from peaceful use of nuclear power?

Though each of these sets of questions represents a focus on different technical domains and social phenomena, each illustrates the interconnectedness of technological and social systems that characterizes our modern world. It is within this intersection of technology and society that Technosocial Predictive Analytics (TPA) [5] exists. The goal of TPA is to "create decision advantage in support of natural decision making through a process of analytical transformation that integrates psychosocial and physical models by leveraging insights from both the social and natural sciences" [6]. In the information security domain, TPA helps the analyst anticipate and counter threats to national security and social well being that originate through this interaction of society and technology. Whether these threats are man-made or natural, malicious or unintended, our ability to create computer models that help us think robustly about plausible future scenarios is increasingly being used to improve our understanding of the consequences emerging from the complex intersection of human society, technology, and the physical environment.

In this paper, we describe the Knowledge Encapsulation Framework (KEF) [7], a platform for managing information, marshaling evidence, empowering collaboration, and automatically discovering relevant data. After discussing related work, our motivation for developing KEF, and its technical implementation, we will explore both the general KEF model for applying the framework and its real-world experiences.

Related work

The underlying research behind KEF is based on research done in a number of domains over a number of years. Experts systems research [8,9] have tried to capture the tacit knowledge residing within a specific domain (usually through the elicitation of that knowledge from subject matter experts [SMEs]) so this information can be shared and transferred to other members [10]. KEF itself does not attempt to master or understand the SMEs' knowledge and evidence as a learning system might. KEF instead focuses on streamlining the research and modeling processes by creating a collaborative environment for SMEs to come together, organize and share information, and provide transparency to help connect research, data, and the types of dialog that occur naturally between researchers. KEF therefore is an environment that allows for the discussion and evolution of new knowledge and ideas and not a more anthropomorphic representation that may appear to have human form and can listen and talk to the user [11].

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 [10]. Finally, typical expert systems focus on a very narrowly defined domain such as Mycin [12] and CADUCEUS [13] (both medical diagnosis systems), NeteXPERT [14] (network operations automation system), KnowledgeBench [15] (new product development applications), and Dipmeter Advisor [16] (oil exploration system). KEF, while similar in many regards to these other examples, is distinctly different as it is specifically designed to be widely applicable to many domains allowing for customization to meet specific domain needs and requirements.

Collaborative problem solving environments (CPSE) are another analogy for this concept. The Pacific Northwest National Laboratory (PNNL) has a long history of building CPSEs for U.S. Department of Energy (DOE) scientists [17], such as the DOE2000 Electronic Notebook Project [18] and Velo [19]. Watson [20] reviewed a number of organizations pursuing CPSEs including other DOE sites (e.g., the Common Component Architecture, Collaboratory Interoperability Framework, and Corridor One Project) as well as the U.S. Department of Defense (e.g., Gateway), NASA (e.g., the Intelligent Synthesis Environment, Collaborative Engineering Environment, and Science Desk) and numerous university efforts (Rutgers University's Distributed System for Collaborative Information Processing and Learning, the University of Michigan's Space Physics and Aeronomy Research Collaboratory, and Stanford's Interactive Workspaces). Shaffer [21], in his position statement on CPSEs, 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 – for example, components to enable three-dimensional molecular visualization for biologists – are most often directly related to the domain.

There are a number of domain-specific applications that a predictive analyst might use. IBM SPSS [22] and SAS Analytics [23] are both marketed towards a business analytics/business intelligence audience and provide capabilities such as text analysis, data mining, visualization, model integration, and statistics. Palantir [24] also markets to business clients, but also has a growing reputation in the intelligence community for being able to mine data from disparate sources (e.g., CIA and FBI databases) and combine them into a single, structured repository. Each of these examples represents widely used predictive analytics applications; however, each is lacking in key areas. Specifically, they do not offer a collaborative framework with capabilities for harvesting content from the internet and automated semantic annotation. They also have not addressed the growing need for being able to combine traditional data repositories with social media data.

Perhaps the most currently available technologies most similar to KEF are "web 2.0" information stores. Examples include encyclopedic resources such as Wikipedia and Knol that rely on the "wisdom of the crowds [25]" 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. Like KEF, some of these systems use tags to provide an informal taxonomy, but the domain scale is typically very wide (in the case of Wikipedia, the goal is to provide an encyclopedia's worth of knowledge). Project Halo [26] is a 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, however. The KEF approach is to share the load between automated information extraction tools and domain experts. 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 provides significant semantic markup and encourages SME engagement.

Motivation for this work

Our work on KEF is motivated by two goals – one specific to the task of TPA, the other more general. The first goal is to provide a framework that meets the specific knowledge management requirements imposed by the multi-disciplinary character of TPA, supporting the ability to:

- collaborate across multiple disciplines
- marshal evidence in support of model design and calibration
- provide transparency into the models being used

Our implementation of features supporting these requirements is discussed in detail in subsequent sections of this paper.

A second, more general goal of this work is to provide a framework that shifts the focus of analysts towards tasks that add value to their data and away from the more mechanical aspects of data collection. It is not uncommon for intelligence analysts (a specific type of knowledge worker with whom the authors have experience) to spend 80% of their time collecting material for their task, thanks in part to the previously mentioned access to publications on the internet, leaving only 20% of time for the analysis [27]. In the research described herein, we aim to address the data quantity 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 collaborative workspace for team members. This framework is also designed to establish provenance, linking data from sources directly to a research area for maximum productivity and pedigree.

Technical implementation

At its core, KEF is a blending of open source software projects and custom development. KEF seamlessly integrates

these separate components into a single environment, providing users with a suite of features and capabilities that no single KEF component can provide on its own.

MediaWiki [28], the same software that powers Wikipedia, forms the foundation of KEF. The wiki provides KEF with many standard web content management system (CMS) features and functionality such as user account management; the ability to easily create, edit, and delete content; a customizable theme engine; attribution of authors for not only the creation of content but all edits and deletions; and perhaps most importantly, a framework for importing community and custom created extensions. As each piece of content is created, MediaWiki creates a new web-based "page" to store its contents. All data from the wiki are stored in a MySQL database. For the author and any subsequent editors, the wiki provides a version control system, ensuring that any subsequent edits, deletions, or moves are preserved for provenance.

Despite being a powerful CMS, these features alone are not sufficient to accomplish the goals set forth by the KEF project. Even though MediaWiki stores its content in a database, each page of content is stored as a single field of text. To a user reading the page, this is acceptable because the user has no direct interaction with the database or underlying functionality. However, for a user who wishes to perform advanced queries across multiple pages, it is less than adequate. Krötzsch et. al. [29] created an extension called Semantic MediaWiki (SMW) that integrates semantic features into the base MediaWiki framework. Extending MediaWiki in this way provided the capability to rapidly sift through the content in the wiki based on the semantically tagged text. KEF uses the Semantic Forms [30] extension to provide manual semantic markup within the wiki pages. Not only does this alleviate the need for a user to learn wiki syntax, a web programming language similar to HTML, but by providing user-friendly forms for data entry, it ensures consistency because semantic properties are applied automatically when wiki pages are created. In addition to properties, each page in the wiki is associated with a template, which controls what information is displayed to the user and how it appears, and a category, which groups similar types of pages together (e.g., all journal articles might be in a "publications" category).

For example, an analyst might have a collection of publications that needs to be tracked with KEF. Some of these publications might be journal articles, books, conference papers, technical articles, technical reports, etc., and as a result, each might have quite different information associated with it. The publication category therefore would be used to group like content together, but each type of publication would have a custom form to capture its information and a custom template to display its information properly.

In a traditional MediaWiki environment, a security analyst could still create a series of pages, each representing a different type of publication in a publication category. The analyst could also perform text-based searches to locate a particular string of text located within one or more of the pages in the wiki. This is how many commercial wikis, such as Wikipedia, function. Within KEF however, that same analyst would have access to much more powerful searching mechanism. Each field that is filled out with the semantic form can be converted into a facet in a faceted browser, [31] a method of filtering and reducing quantities of information, to rapidly filter the collection of publications to a more manageable subset based on a selection of semantic properties. Instead of the traditional "search results" page, the page would be a dynamically updating one where the analyst has the ability to drill down into the content and more easily find relevant information.

For example, as seen in Figures 1, 2 and 3, the analyst would be presented with a set of results in the faceted browser. From here, the analyst may select a particular author, publication date, or interesting phrase to explore the results in a manageable way. If the analyst selected the date "1995-01-01" in Figure 3, all but 2 of the original 145 results would be filtered out. Any of the metadata collected during content entry may be exposed as a facet, giving a high degree of customization to these interfaces and allowing them to be molded to most accurately represent the content to be explored.

KEF blends community and custom extensions to facilitate this faceted browsing capability. Exhibit [18], a

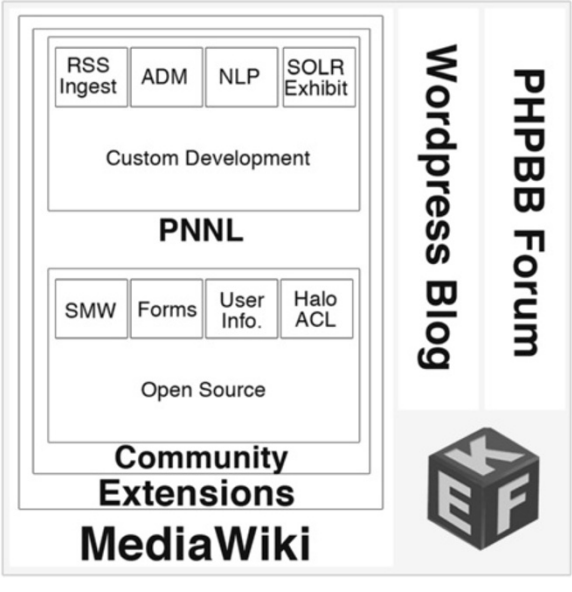

Figure 1 KEF technology diagram. This figure illustrates the various components that have been brought together to create KEF.

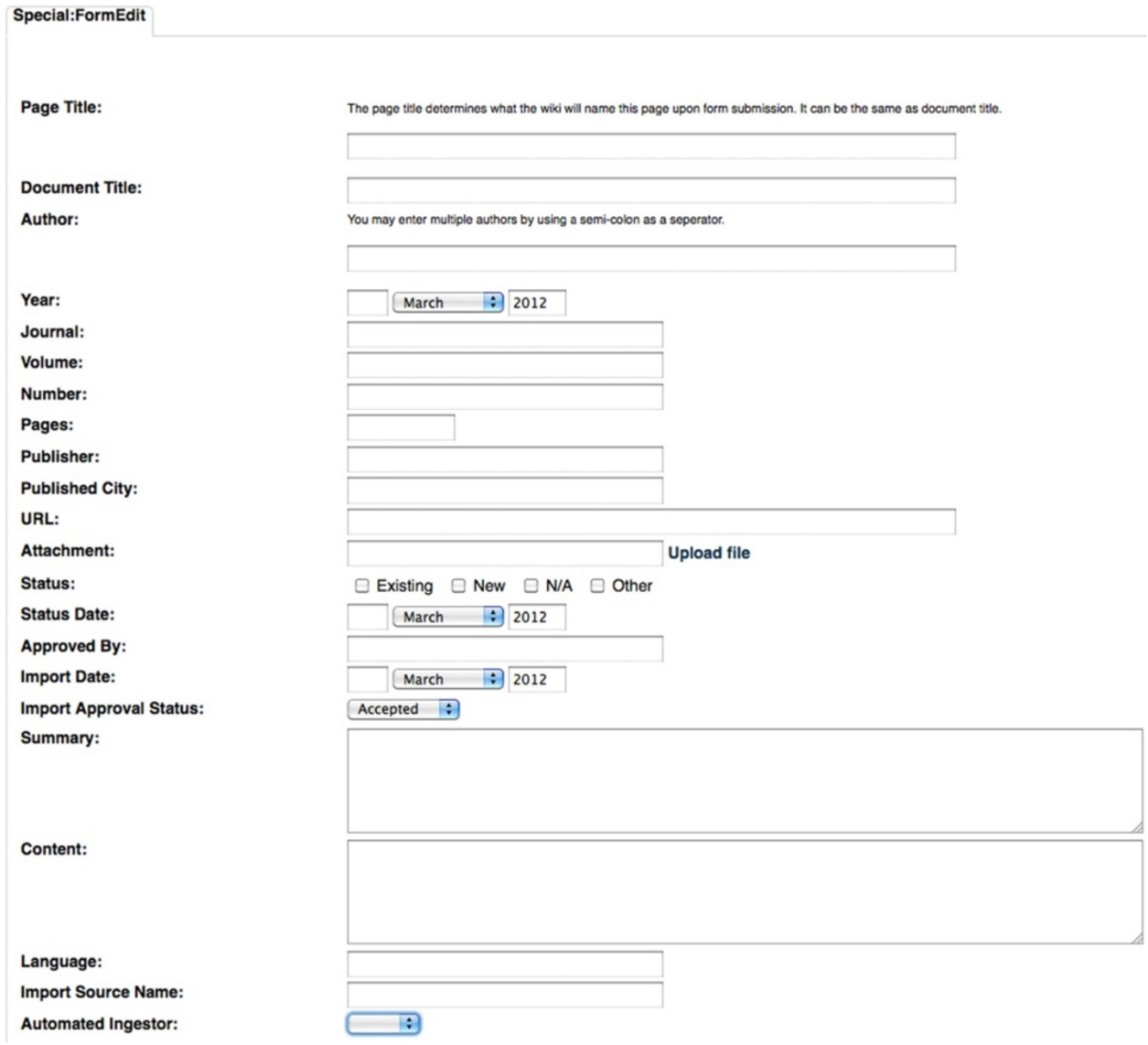

Figure 2 Journal article entry form. This screenshot illustrates the form that aids the analyst when creating a journal article.

product of the SIMILE project at the Massachusetts Institute of Technology (MIT), provides a number of simple visualizations for the semantic data such as a Timelines, Table, Map (powered by Google Maps), and Calendar. The value of these visualizations is amplified by the faceted browsing technique, allowing the user to remove any of the pages that do not match their filters. KEF updates the visualization with each new selection, reducing the amount of data that the user must actively view. The KEF development team has integrated the research done at MIT on the Exhibit project with Apache SOLR [32] technology to significantly improve the scaling of Exhibit, giving the user instantaneous access to the data contained in their KEF site, even when there are tens of thousands of pages in the wiki.

Beyond providing basic content and user management, KEF serves as a collaborative environment, fostering discussion and the sharing of information. Several enhancements are necessary to facilitate this capability in the wiki. We have introduced the concept of User Profiles into KEF through a customized version of the community extension Social Profile [33]. These profiles might contain the standard "social networking" type of information such as name, email address, interests, skills, etc. They also commonly include research interests, publications, projects, and other information that might not be shared on a traditional social networking site (e.g., Facebook, LinkedIN), but would still be of interest to internal collaborators. We have also bolstered the security system within the wiki environment to adequately protect the users and their

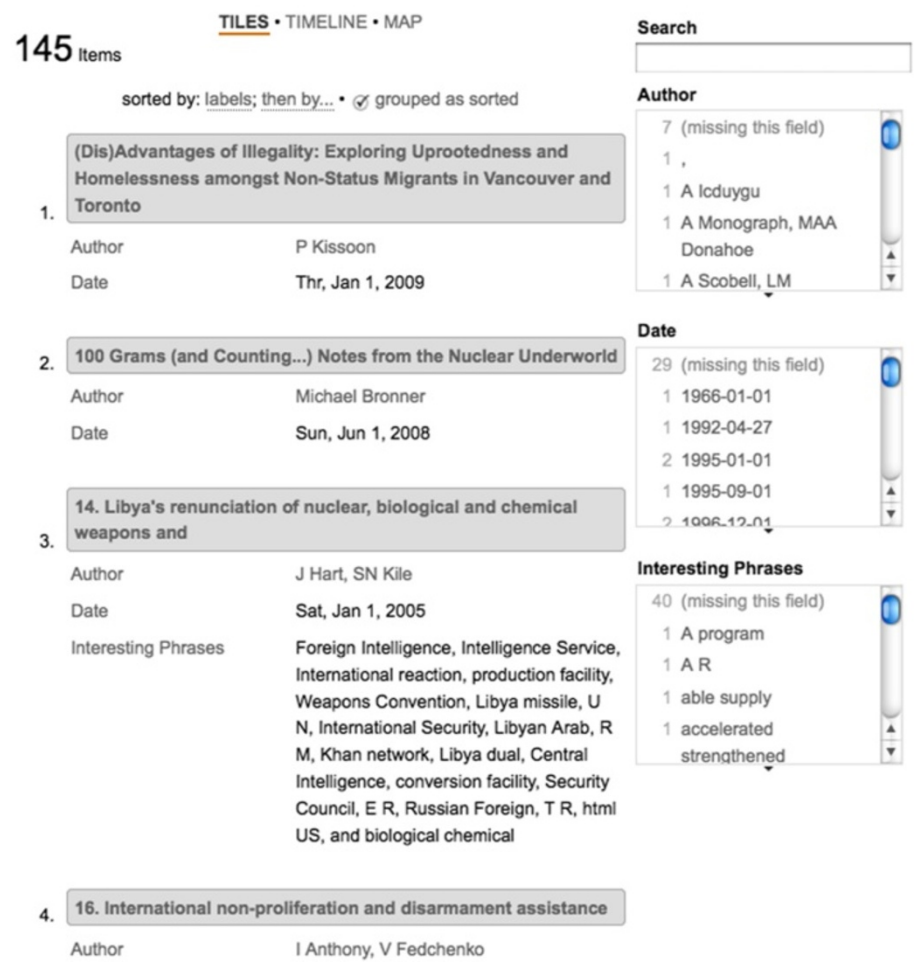

Figure 3 Faceted browser. This screenshot illustrates the faceted browser used by an analyst.

research. A wiki, at its core, is an open collaborative environment. Because of the sensitive nature of some analysts' research, it is often necessary to provide safeguards and access restrictions on their KEF installations. The community extension HaloACL [34] was integrated into KEF to provide security for these cases. This extension provides the capability of hiding complete pages and sub-page elements from users outside a particular user class. For example, a team of analysts might be spread across several institutions, requiring that their KEF installation live on a publicly available web server. While a brief welcome screen and general explanation of the project might be available for public consumption, none of the research, discussion, or modeling that goes on within KEF should be available publicly. Through HaloACL, that installation can be secured so that only registered and approved users that belong to the team of analysts can view or edit the sensitive data in KEF. On some installations, that is all of the data while others only protect a small subset.

Other community extensions add functionality such as the ability to construct widgets for commonly used code (e.g., embedding social video such as YouTube), add new semantic views for data (e.g., a sortable, printable, color-coded spreadsheet), use simple programmatic functions in wiki markup (e.g., if statements and arrays), a What You See is What You Get (WYSIWYG) editor, etc. In addition to custom development already highlighted in this section, the KEF team has created a significant number of extensions to facilitate specific functionality. These extensions will be covered in more detail in the KEF Model section below.

KEF relies heavily on MediaWiki for content management, but MediaWiki is not the only component within the framework. PHPbb [35] (PHP Bulletin Board) is a web-based forum application. While an analyst could easily share links to content in the wiki through an email or instant messaging client, these solutions are often lacking when it comes to recalling

the conversation in the future or sharing it with other collaborators. A discussion forum provides a central resource that anyone with the appropriate access can view, engage, and share. As we will outline, the ability to discuss the activities in the wiki with other members of a research team to solicit feedback and knowledge sharing is critical to the success of the KEF model. KEF also uses Wordpress [36], a web-based blogging engine. A wiki is designed to have infinite layers of content, while a blog is designed to give users a chronological view of new information. Many KEF installations use the "blog" feature as an announcements or tasking platform to disseminate changes or new information rapidly among the user community. The KEF Model

Figure 4 illustrates the collaborative process followed by a team using KEF. Each numbered section represents a collaborative effort that is part of the KEF process. The specific example in Figure 4 focuses on a demonstration of KEF's capabilities that studies nuclear proliferation and illicit trafficking. This model focuses not on managing existing evidence (although KEF can play that role), but on discovering new evidence, fostering the collaboration of SMEs, aligning evidence with data models, and analyzing these data and evidence.

At its simplest, this process is:

1. Set Up Environment
2. Automated Discovery Process
3. Evidence Marshaling
4. Evidence-Model Alignment
5. Analytical Gaming / Analysis

Throughout each stage in this process, KEF steps outside of the standard wiki model to fuel collaboration. To that extent, KEF's discussion forum allows team members to flesh out ideas and participate in threaded conversations. Team members are automatically notified via email when new topics or replies are posted, ensuring that busy team members are kept in the loop even if they do not view the forum regularly. KEF's blog allows a team member to reach out to the rest of the team and alert them to new content in the wiki or a new discussion in the forum. Team members receive announcements from the blog either as a subscribed RSS feed or via email.

Stage 1: Set up environment

When starting a new instantiation of the environment, the KEF team meets with the SMEs and modelers to understand more about the domain (Stage 1 in Figure 4).

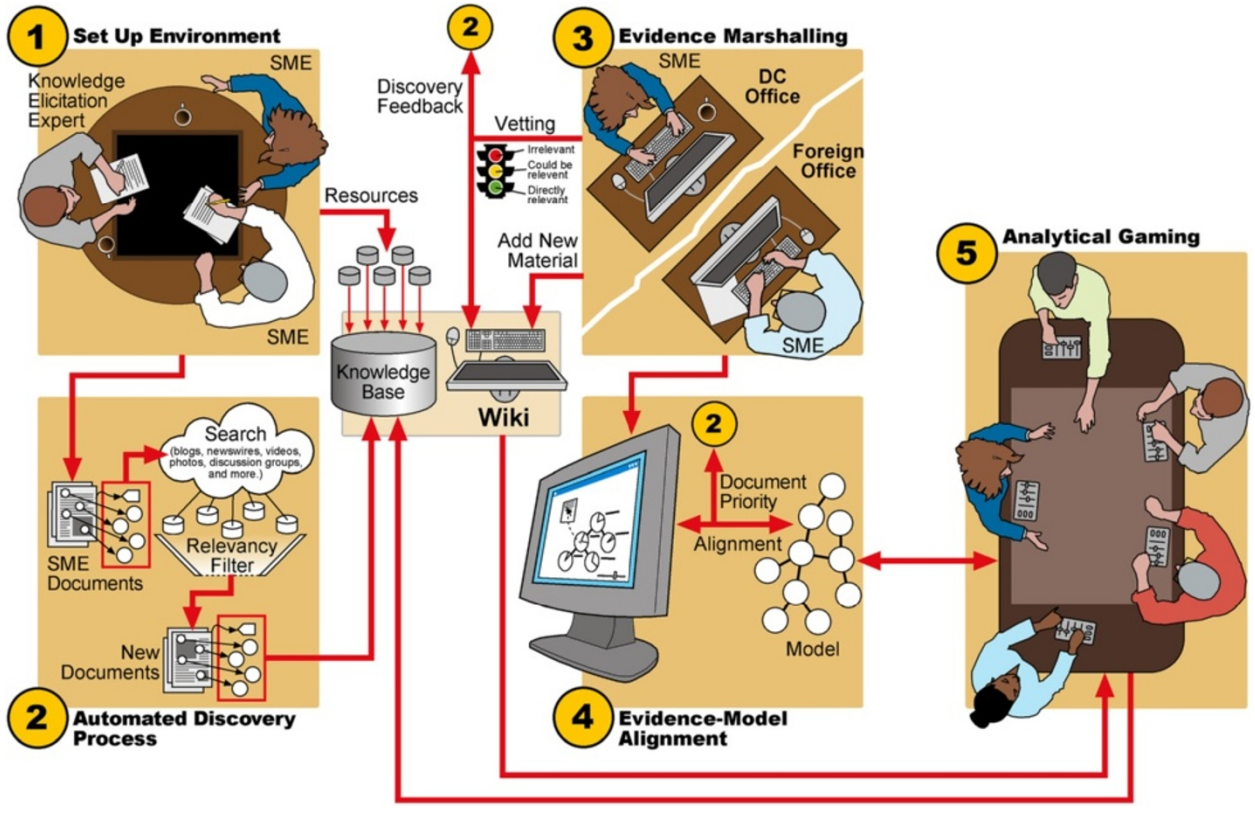

Figure 4 The KEF model. This figure shows a detailed illustration of how the KEF process takes place.

The KEF team is composed of computer scientists, designers, and developers. As a result, projects rarely are in a domain that matches the team's experience. These initial meetings with SMEs provide critical insight into the problem space and the desired outcome of the project. Key resources such as data sets, specific databases, important documents (e.g., journal articles, technical reports, etc.), and social media sources (e.g., specific users, topics, or sources) are gathered as the team seeks an understanding of the major domain concepts.

Based on the requirements outlined by the SMEs and modelers, KEF is customized based on project-specific requirements to easily incorporate the project's key resources. A typical KEF environment is deployed to a web server for development using a custom continuous build system. As the members of the research and KEF teams work on the site, changes in content (which are primarily stored in the MySQL database) and changes in the underlying framework (which are primarily stored on the web server) can be made simultaneously, allowing for the rapid development process that is often necessary in a research environment.

Once the environment is set up, the KEF team usually creates a series of Semantic Forms for the SME to use for manually entering content. As this new content is added to the wiki, the underlying semantics = mark up the text. Semantic Forms itself allows for basic markup, using each of the different fields in the form to represent a semantic property once the page has been created. KEF has also introduced a series of custom Natural Language Processing (NLP) tools that search through the submitted text adding additional annotations. The goal of these annotations to the unstructured data is to assist the SME in adding additional structure to the unstructured data that will support discovery and alignment of evidence. The current set of automated annotations includes:

- named entity mentions
- automated categorization
- statistically improbably phrases
- sentiment analysis
- event recognition

Named entity mentions annotations are used to indicate where entities of certain types of interest are referred to in a document. For recognizing named entities such as proper names, dates, times, and locations, we use two approaches: a statistical approach to provide coverage for general entity types (e.g., Person, Location, and Organization) and a dictionary-based approach to provide precision for domain-specific types. KEF's statistical named entity recognizer (NER) annotator can also use the Stanford NER tagger to tag people, organizations, and locations based on a linear chain Conditional Random Field sequence classifier. The dictionary-based NER annotator uses lists of terms provided by the SME to tag entities relevant to the particular domain, such as specific types of people, organizations, or technologies.

Automated categorization annotations identify documents belonging to particular categories. The process for determining these categories starts with the SME providing some example documents. A maximum likelihood estimator (provided by LingPipe [37]) is trained on these categories and documents. As new documents are added, KEF can automatically place them into the appropriate category.

Statistically improbable phrase annotations identify phrases in a document that are deemed unlikely as compared to some background corpus. The sort of phrases identified can vary based on the background corpus used. For example, a general-purpose background corpus is used with a similar approach for Amazon's Statistically Improbable Phrases [38] to produce domain- or topic-specific terms in books. Similarly, a same-domain corpus of earlier documents can identify emerging themes and terms over time as used by Google News and similar tools. The functionality of our statistically improbable phrases annotator is based on LingPipe. As a background corpus for each document we use a collection of public domain novels, providing a generic model that allows topic-relevant terminology to emerge.

Sentiment analysis is performed to identify polarity (positive or negative) of documents or passages. This is driven by lexicons, which may be customized for specific domains. These annotations can be used for searching for evidence supporting specific opinions.

Event recognition is used to automatically annotate mentions of events of interest and the entities that that have roles in the events. Events and entities are identified using an information extraction pipeline and labeled according to types defined in an ontology. Event ontologies can be centered around domains – such as terrorism or technology [39] – or types of evidence – such as rhetoric [40].

With the structure given by these automated annotations, features of the semantic wiki such as the faceted search and summary views can be used by the SME to home in on specific content or pieces of content of interest for identifying evidence. For example, to identify the current state of networks of interest, the SME can search for mentions of entity types representing people of interest (e.g., Denied Person).

A threaded discussion forum and blog are also often deployed with the wiki in the earliest stages of the KEF development cycle. The forum will house discussions related to the models and data being gathered within the wiki. We begin with the forum in place to ensure that as content is manually entered in Stage 1, automatically harvested in Stage 2, integrated with the wiki in Stage 3,

and aligned with data models in Stage 4, the SMEs will have a consistent area for holding collaborative discussions. Even during the analysis stage, a SME can return to the same discussion space to resume a discussion from a previous portion of the project. We also use the blog to highlight new features, or pieces of content, that might be of interest to other members of the team.

At the end of Stage 1, KEF has a functional blog, wiki, and forum and is available for the research team to begin collaborating, although at this time the amount of content is limited to only those documents manually entered. For example, this could include those reports and other documents considered to be excellent examples of the types of information the SME's and modelers hope to use to drive their models. In addition, this could also include structured data sets.

Stage 2: Automated discovery process

In Stage 2, we introduce the automated discovery mechanism (ADM). This suite of tools enables the SME to use the content entered during Stage 1 to automatically locate content on the internet (and other, potentially secure or otherwise restricted data sources) that may be statistically relevant. Through the semantic markup that was done as these "seed documents" were created, the ADM captures the essence of those documents (e.g., named entity mentions, automated categorization, statistically improbably phrases, sentiment analysis, event recognition) and searches across other data sources to identify potentially relevant material, covering both traditional and social media, such as:

- Google Scholar
- Opensource.gov
- CNS Nonproliferation Databases
- Microblogs (e.g., Twitter)
- Blogs dedicated to nuclear nonproliferation discussions
- The Nuclear Suppliers Group Trigger and Dual Use list
- The U.S. munitions list (category I-IV)

For each document identified, a relevancy metric (specifically, binary term occurrence) is computed to evaluate whether the document is truly related and not just a copy of the same document or too distinct to be useful. A researcher interested in expanding the search into new domains or topics of interest can add additional seed documents to KEF, which will in turn cause the ADM to expand its search to include those new concepts. Each document discovered through the ADM is harvested into KEF, passing through the same NLP tools as content that was manually entered and stored in a temporary repository while it awaits review by a SME.

Stage 3: Evidence marshaling

In Stage 3, users make use of KEF's faceted interface and summaries to browse the harvested content from Stage 2 and manually vet each piece of content, allowing the SMEs to decide which pieces of evidence should be introduced into the wiki. We recognize that no matter how thorough our NLP tools are, an automated harvesting process will inevitably find data that are not of interest to the SMEs. The goal here is that their vetting decisions can be fed back to the discovery mechanism to help improve the quality of the ADM process.

The faceted interface will load a series of documents that the ADM has harvested. A summary of information (e.g., source, title, categories, named entities, etc.) will be shown to the SME, and a series of facets will be available with similar content. The SME can rapidly go through the content and mark which documents should be accepted into KEF and which should be deleted. At any time, SMEs continue to add their own material into the environment adding to the vetted documents being harvested by the ADM.

Stage 4: Evidence-model alignment

In Stage 4, users can upload their model structure to the wiki and the environment will parse the structure and associated properties. Currently, this feature is in place for Bayesian Analysis of Competing Hypotheses (BACH) [41] models that are represented in XML, but similar visualizations can be added for other types of models. The user can then select specific parts of documents to connect to parts of the model (e.g., a paragraph of a known nuclear trafficking suspect entering the country could be aligned with the model node entitled "Suspect Geographically Linked to Target"). With large numbers of users, the goal is that the system will start automatically classifying the textual annotations linked to a particular node. These can be used to recommend other documents that the user should examine within the wiki and prioritize incoming material. This is especially important when the corpus of discovered documents is large. Specifically, the system attempts to classify the piece of text that users align with model nodes in order to characterize their linguistic structure so that it can try to identify this signature elsewhere.

This approach is based on research done by Y. Li et al. [42] in which two sentences are semantically compared using the WordNet [43] ontology, weighted by the frequency of the words in a large corpus and further combined with the similarity of word order among the sentences. We have expanded their research by preprocessing the sentences in order to semantically compare the words in those sentences. The first preprocessing step is to tag the various parts of speech (nouns, verbs, adjectives, etc.). The next step is to disambiguate the word

sense of each of the words as outlined by Kolhatkar [44] so that "blue" in "I'm feeling blue" and that found in "The sky is blue" are considered separate with distinct meanings. The third step is to do a lookup to convert the form of the words to that found in WordNet. All words whose parts of speech or word sense cannot be determined, as well as those not found in WordNet, are removed before the sentence comparison is attempted.

The results of our research are promising. For example, the sentence "The threat that terrorists could acquire and use a nuclear weapon in a major U.S. city is real and urgent" was compared against a document containing 141 sentences. The most similar sentence retrieved was worded: "A dangerous gap remains between the urgency of the threat of nuclear terrorism and the scope and pace of the U.S. and world response." The first sentence talks about a real and urgent threat of nuclear terrorism while the second suggests that the international community's pace to respond to that threat is insufficient compared to the urgency of the threat. The next step is to expand from sentences to paragraphs while maintaining the level of accuracy experienced at the sentence level [45].

We are currently researching novel methods to expand this concept to dynamically assign evidence as new content is fed into the corpus as well as learning from the analyst's responses whether or not the evidence that the automated process finds is relevant. The algorithm will then use this feedback to further improve the discovery mechanism.

Stage 5: Analytical gaming/Analysis

In Stage 5, the model, fully parameterized with attached evidence, can be exported from the wiki environment and used within a separate analytical tool suite. Figure 4 shows a serious game that enables decision- and policy-makers to perform "what if?" analysis. The game can reach back into the KEF environment to utilize real content and push results and decisions back into KEF for retrospective analysis. KEF can also be used to inject material directly into a running game in order to change the focus and bring the game back on track. Figures 5

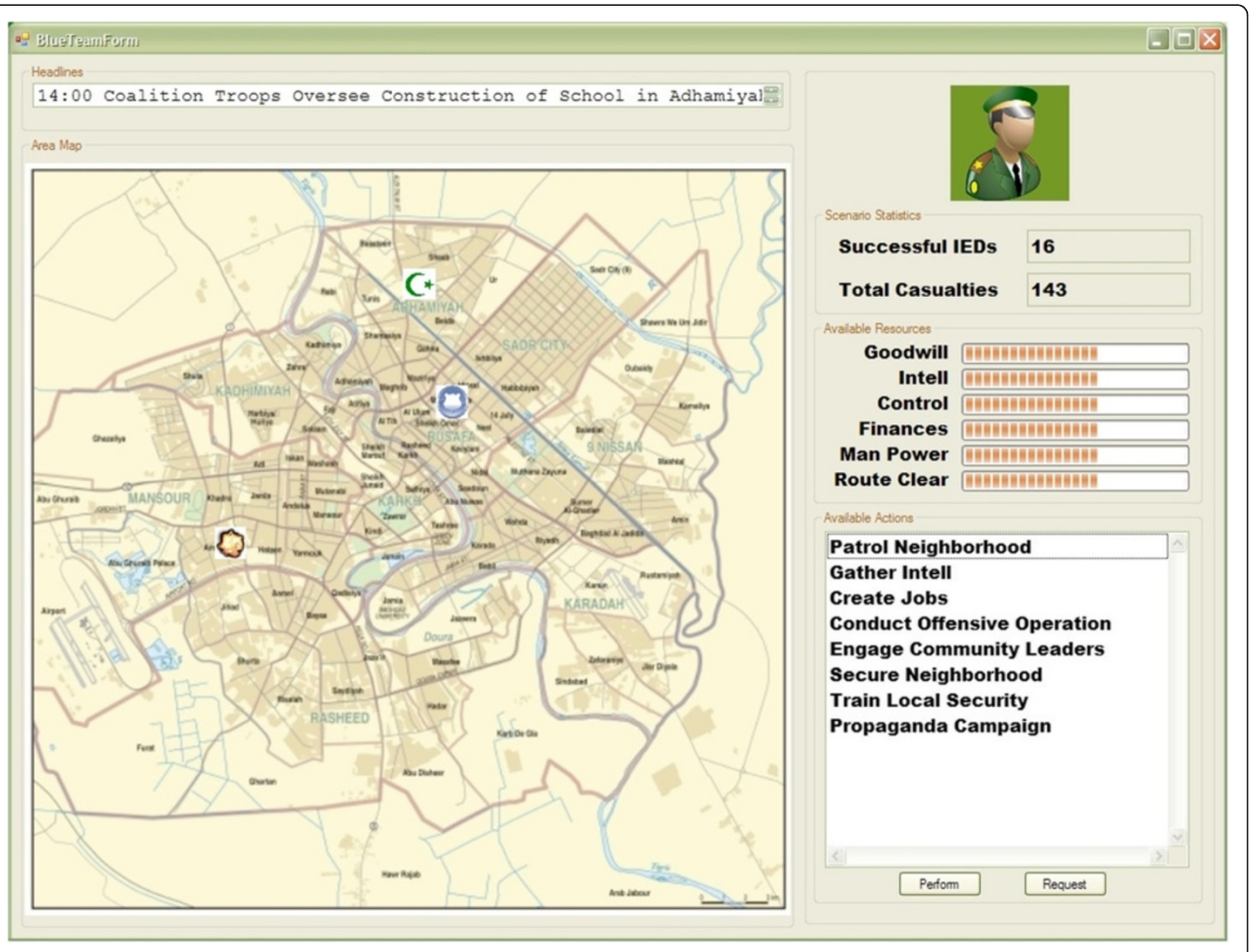

Figure 5 The IED Game. This screenshot shows a portion of the Analytical Game used in the Illicit Trafficking Demonstration (ITD), which will be discussed in the next section.

and 6 show some examples of how this linkage has been exercised with an Improvised Explosive Devices (IED) serious game [46] and an Energy Infrastructure Security serious game [47].

KEF also includes a basic chart and graph API, allowing users to visualize data sets metadata about evidence (e.g., comparing the number of pieces of evidence from a series of categories).

Case Study: The illicit trafficking demonstration

The Illicit Trafficking Demonstration (ITD) was intended as a showcase of the capacities provided by KEF, particularly its integration with the BACH and Analytical Gaming [48] frameworks. The demonstration showcases KEF's handling of the interaction between SMEs, the documents they have entered into KEF, and the analytical models built from those documents. The end goal of the demonstration was to present a cohesive environment where SMEs, analysts, and other interested parties could collaborate on the construction and execution of a particular analytical model within a single environment. We describe the implementation of KEF within PNNL's Technosocial Predictive Analytics Initiative (TPAI) [5] capstone demonstration below.

ITD was constructed for analysts working in the domain of nuclear trafficking and nonproliferation. As a regular part of their job, these analysts are often asked to research the formation of illicit nuclear trafficking networks, how nuclear materials might move and proliferate through those networks, and the relative likelihood that particular countries or political actors might engage in nuclear trafficking activities. A possible outcome of this research is a model describing the likelihood that a particular type of nuclear material might be transported into the United States. During our research, we found that many of the analysts we interacted with were overwhelmed by the amount of data they could interact with in their current toolset [6]. These data primarily comprised web search results obtained through a variety of sources.

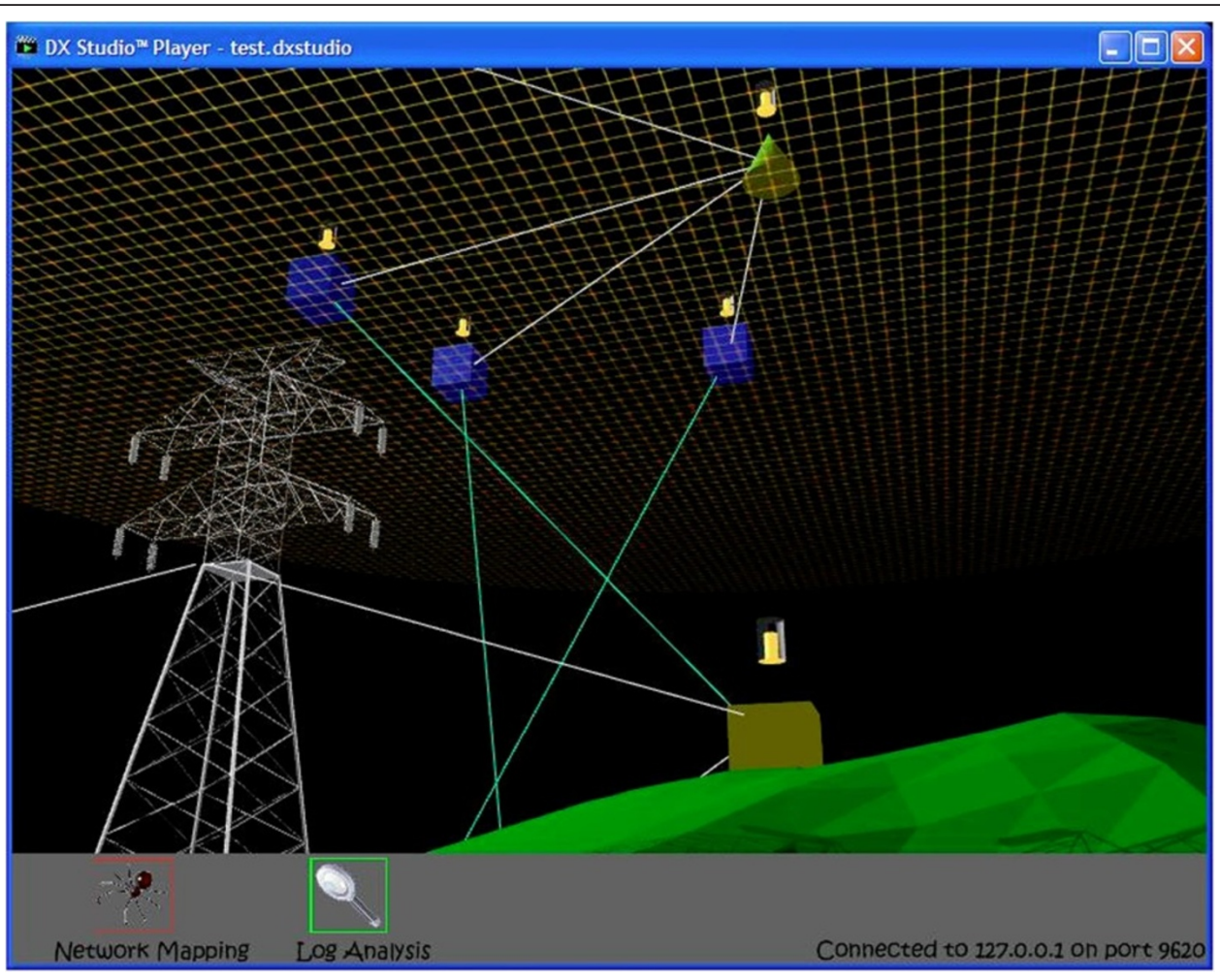

Figure 6 The Energy Infrastructure Security Game. This screenshot shows a portion of the Energy Infrastructure Security Game.

To help offset this overload, the KEF Model (specifically Stages 2 and 3) is formulated to streamline the information capture and analysis process.

When analysts visit the ITD site, they can see at a glance the various types of content that have already been harvested and incorporated into KEF. In Figure 7, the majority of the data has come from social media sources, primarily blogs. There is, however, also a series of traditional journal articles, books, conference papers, and presentations, some of which were entered manually as seed documents to aid in the ADM. Analysts can select a particular type of document (e.g., social media) and load a faceted browser, as seen in Figure 8, to drill down into that data.

All of these social media data have been automatically harvested by KEF. A SME has already vetted these data and allowed it to be entered into the wiki. After using the faceted browser to narrow down the original results (in

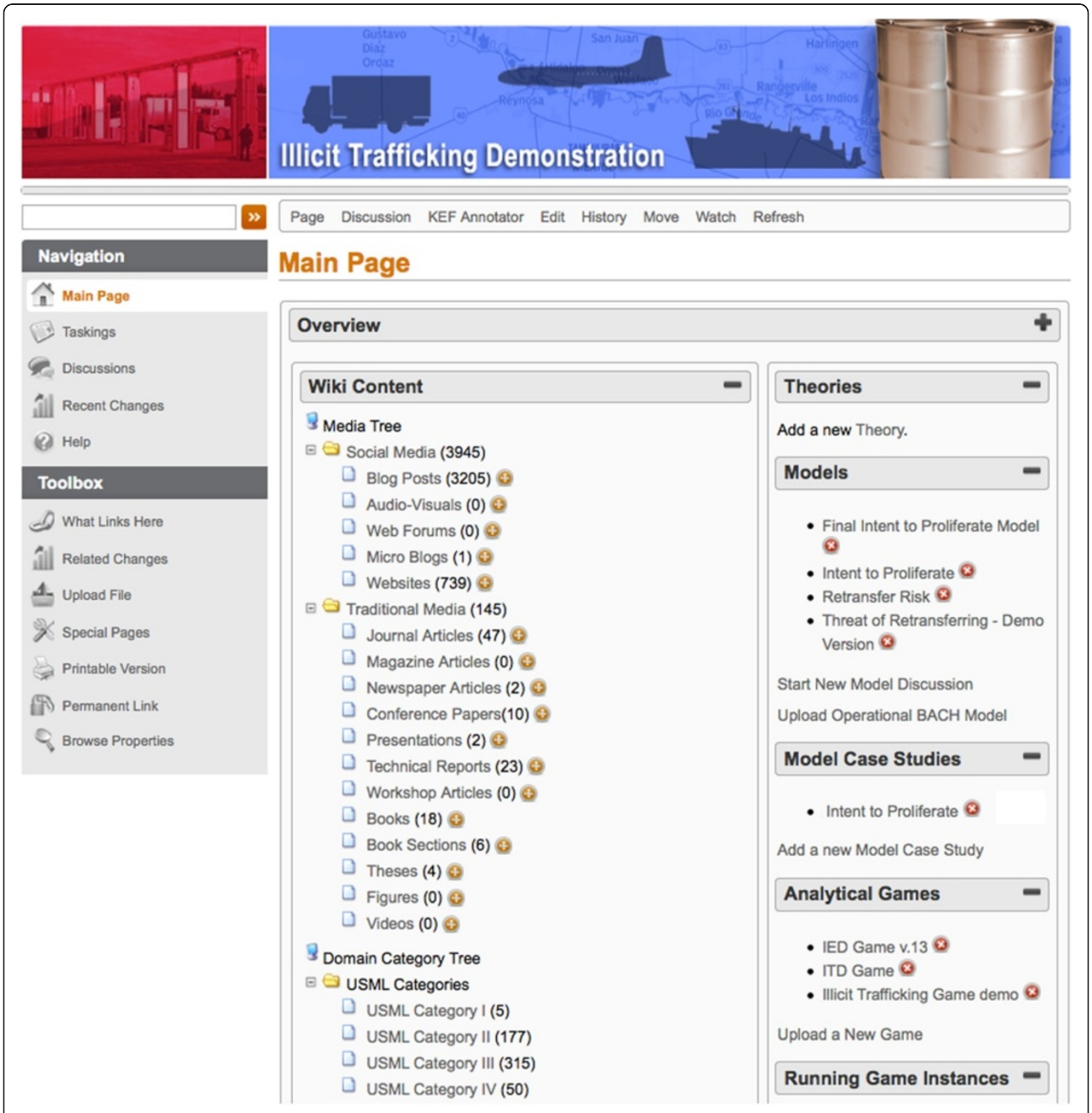

Figure 7 ITD Homepage. This screenshot shows the landing page that a SME would see when visiting the ITD KEF site, including links to the various types of documents, available models, analytical games, and the collaborative discussion forum and taskings areas.

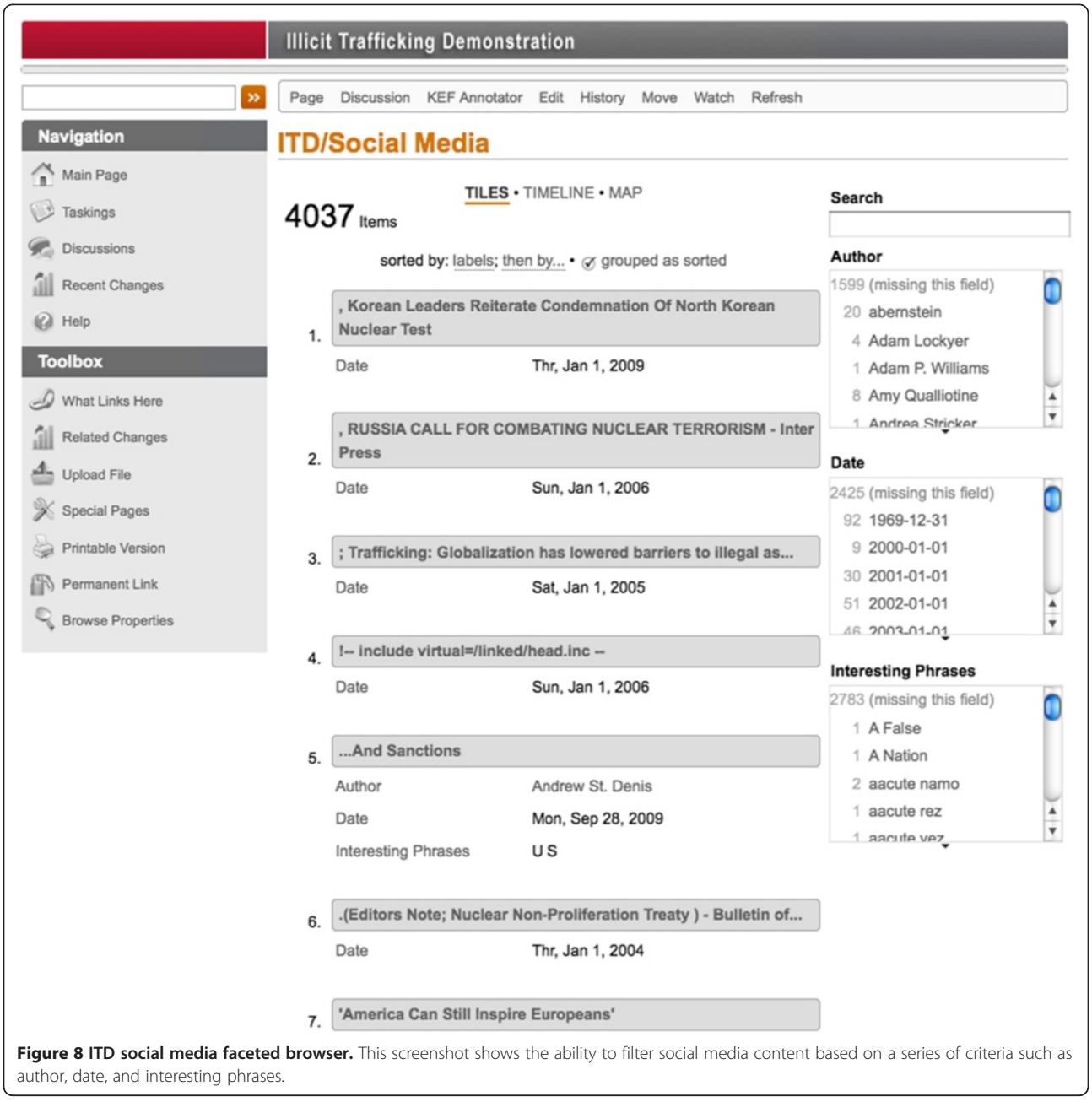

Figure 8 ITD social media faceted browser. This screenshot shows the ability to filter social media content based on a series of criteria such as author, date, and interesting phrases.

this case, we started with approximately 4000 documents), the SME can select a single entry to view the document, as seen in Figure 9, as it was harvested from the web.

This example of an individual record in KEF contain several key pieces of information and capabilities:

- key information about the article (e.g., title, author, source)
- metadata about the article (e.g., KEF project, number of comments, and whether this document is used to seed future harvests from the ADM)
- the original content
- history of the article within KEF (to preserve provenance of harvest, edits, and modifications)

From this individual blog post, the SME has several options. The KEF Annotator (as seen in Figure 10), was custom developed by the KEF team to enable analysts to further semantically mark up the document by highlighting phrases that are of particular interest. The discussion option will transport the user to the threaded discussion area of KEF, as seen in Figure 11, where the SME can join an existing discussion surrounding this particular article or start a new one.

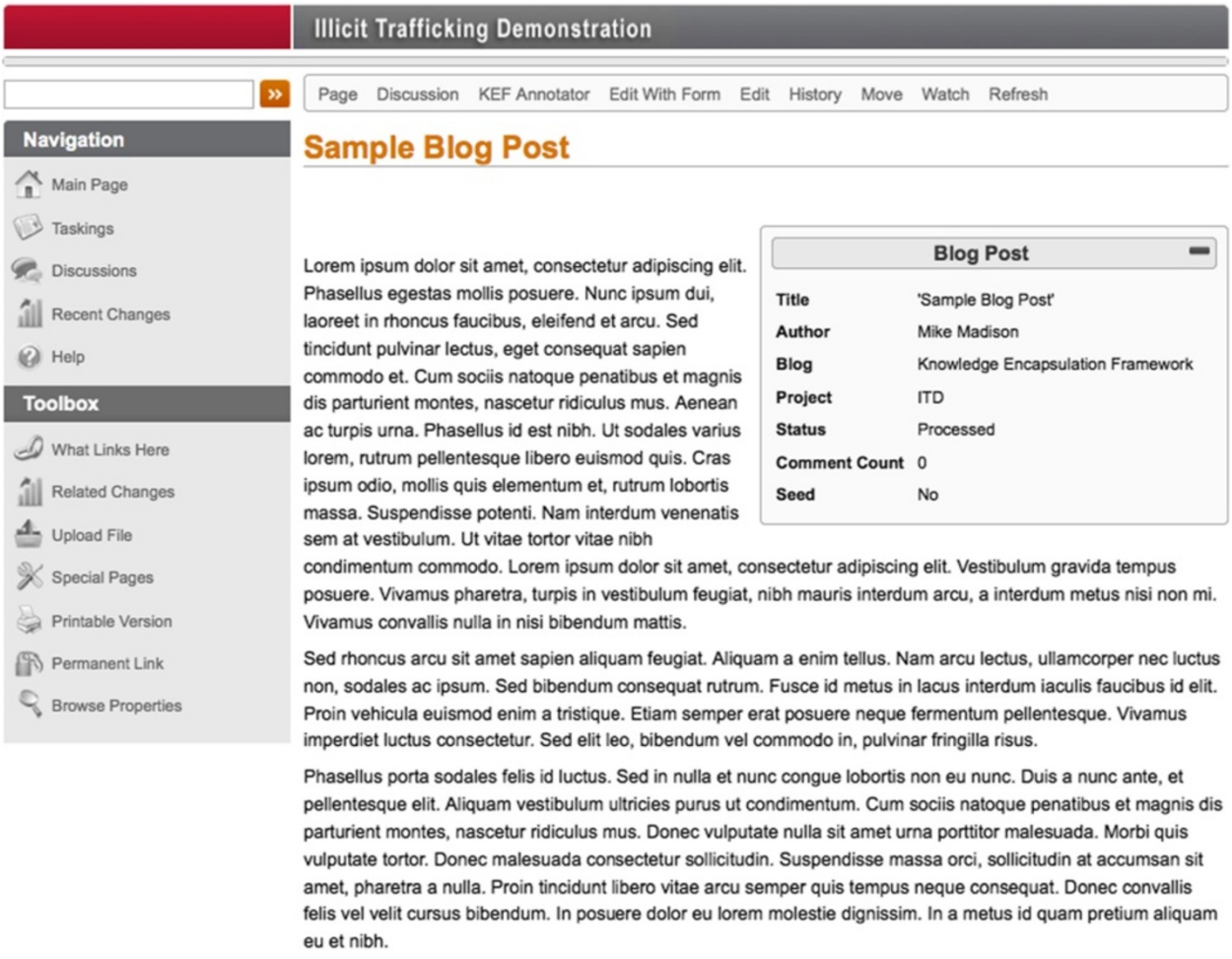

Figure 9 ITD blog post. This screenshot shows harvested content from a blog after it has been placed into the KEF environment.

If the analyst instead decides to focus on an existing model, they could go into a data model directly from the homepage. KEF is not, itself, a modeling framework. However, we have built the capability into KEF to visualize models and associate evidence with particular nodes from the model. Figure 12 shows a BACH model being viewed in the model visualizer.

This model visualizer is accompanied by pages of documentation that explain each node. KEF also allows for the alignment of evidence, both manually entered and harvested through the ADM, with any part of the model.

KEF evolution since the illicit trafficking demonstration

ITD was very precisely targeted at one domain, nuclear trafficking and nonproliferation. Since KEF's inception we have completed projects in a number of other domains such as cyber security [49], renewable energy [50], biomedical nanotechnology, signature discovery [51], multiscale science, semantic technologies, carbon sequestration [52], microbial communities [53], visual analytics, mass spectrometry, and computer supported cooperative work (CSCW). We have found, over the past three years, that our original concept of "what KEF is" to an analyst has evolved somewhat. We find that the overarching concept of KEF, that collection of open source applications, community extensions, and custom development, is largely the same. However, we have also found that KEF implementations can be successful even when they omit some elements of the framework, depending on the needs of the project and the subject domain. These needs have also driven the development of new options in the framework, including the internal management of modeling data and the integration discussion topics in-line with wiki content.

Management of data for models

Multiple projects, across a number of domains, have recently approached us about KEF's data modeling capability. In the past, it was assumed that KEF would be used

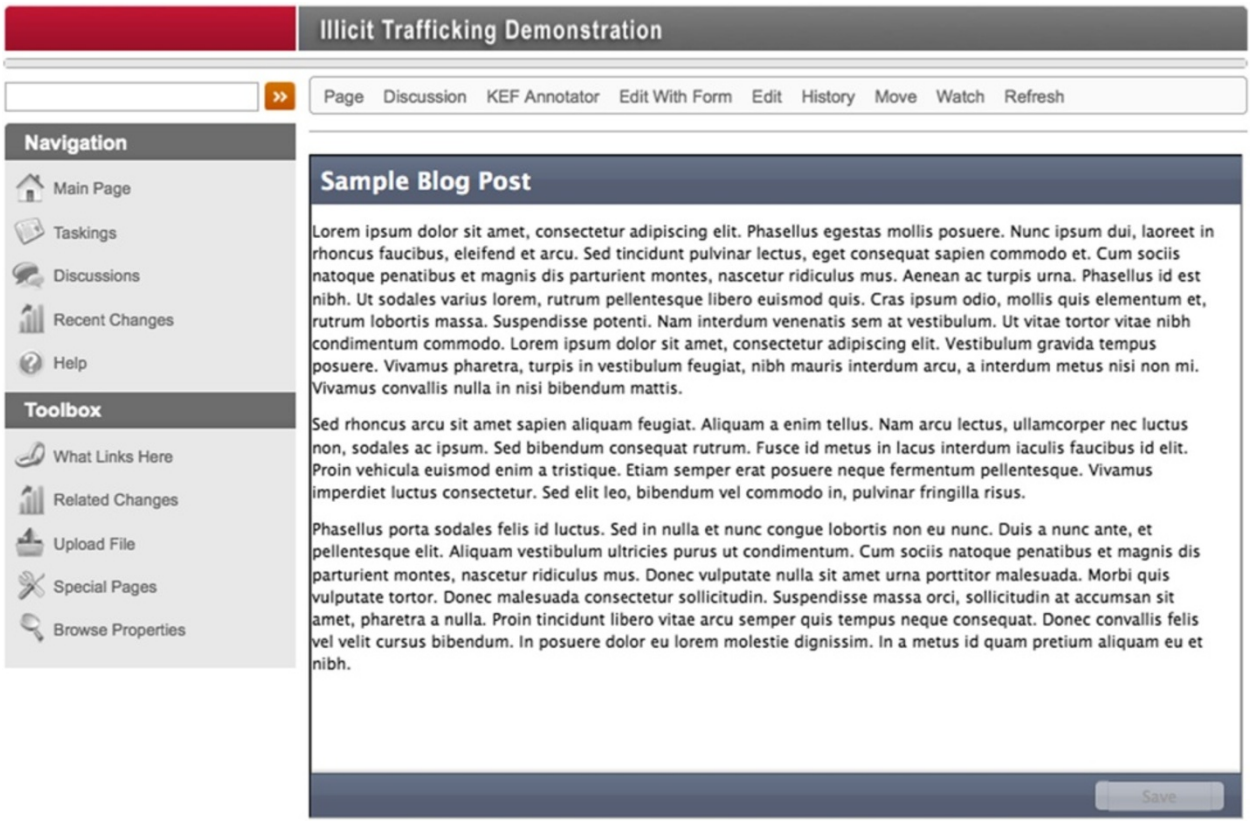

Figure 10 ITD KEF Annotator. This screenshot shows the KEF Annotator.

as a way of marshaling evidence to support models, but this modeling would not necessarily be done in the KEF environment. Current development is aimed not only at aligning evidence with these models (e.g., BACH) but also managing the data itself:

- Recent work with projects dealing with radical rhetoric [54], business intelligence, and building component data has modified analyst workflows. For example, through KEF, an analyst can start with a large data set, use the faceted browser to filter out unwanted or unneeded data, and then send that data directly into a modeling framework running in parallel with the wiki environment. Users can control the data that they run through models much more accurately, thanks to the filtering they apply in KEF before exporting the data into the model.
- Similar work is being pursued to allow data maintained in KEF to be exported directly into visualization tools such as Scalable Reasoning System (SRS) [55] or IN-SPIRE™ [56], giving users access to additional analytical tools beyond the existing gaming and charting frameworks.
- KEF's growing capabilities for managing large quantities of structured data in a user-friendly way give SMEs and other users easy access to data that they might not otherwise locate.
- This increasing use of KEF in managing model data also highlights the benefit of working with SMW, as the wiki already has an established programming language for interacting with its data and allowing external applications to gain access to it.

Discussion threads in-line with content

It is uncommon, unfortunately, for users of KEF to take full advantage of the current implementation of discussion forums. Web analytical data from KEF sites and responses from project managers after deployments often indicate that a discussion forum topic about content in the wiki will see significantly less traffic than the wiki content page itself. The responses to the discussion topic are typically even fewer than the number of "reads" that the topic receives. In an effort to better engage users in meaningful discussions, we developed an in-line discussion feature that allows content contained on a wiki page to be discussed in an integrated, threaded discussion located on the same wiki page. We believe that the threaded discussion forum still has value, as it gives users a view of "what has been discussed since I was last here." We also believe, based on our analytical data and

View unanswered posts | View active topics — View new posts | View your posts

Board Index — All times are UTC - 8 hours [DST]

[Moderator Control Panel]

Mark all forums read

Forum	Topics	Posts	Last post
ITD Model Discussions	13	24	Wed Oct 20, 2010 2:41 pm Roderick Riensche
Other Model Discussions	0	0	No posts
ITD Scenario Discussions	0	0	No posts
Gaming Discussions	0	0	No posts
KEF Discussions	15	52	Wed Aug 18, 2010 5:05 pm Andrew Piatt
Initiative Discussions	0	0	No posts
ITD Feedback	5	8	Thu May 06, 2010 8:49 am Andrew Piatt

Delete all board cookies | The team

Who is online

In total there is **1** user online :: 1 registered, 0 hidden and 0 guests (based on users active over the past 5 minutes)

Most users ever online was **5** on Fri Jul 30, 2010 8:25 am

Registered users: Michael Madison

Legend :: Administrators, Global moderators

Statistics

Total posts **84** | Total topics **32** | Total members **29** | Our newest member **Mikhail Akopov**

Figure 11 ITD discussion forums. This screenshot shows the discussion forums.

Model Image [edit]

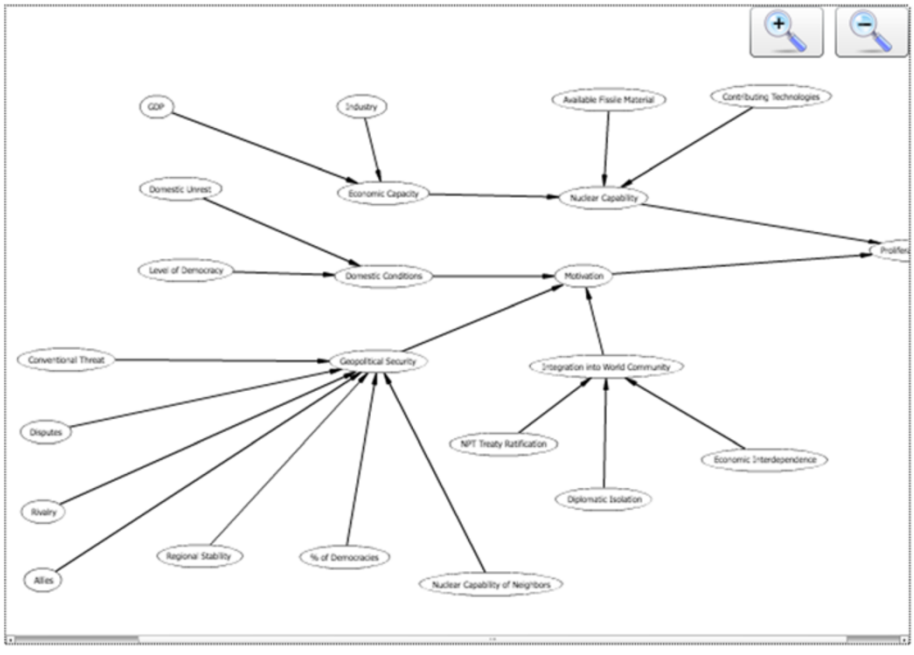

Figure 12 ITD model visualizer. A representation of a BACH model within KEF.

user interactions, that by placing the discussion in-line with the content, more users will be exposed to the conversation and encouraged to participate.

Increased data scaling

In the original KEF model, we expected that analysts would be harvesting a combination of traditional and social media. However, some years ago, the volume of social media data was significantly less than it is at the time of this writing. Clients also have become interested in exploring large data sets (tens of thousands of records) within KEF, using its faceted browser and visualization tools to explore these data. As a result, KEF regularly must increase the scale of the data we can handle. Additional work with Apache SOLR, as well as continued tweaking of the MySQL database, is continually underway to allow for increasing quantities of data to be hosted seamlessly within the KEF environment.

Conclusions

The Knowledge Encapsulation Framework represents a leap forward in the collaborative process of teams across many domains. We have presented a collaborative workspace for analysts to gather, automatically discover, 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 the faceted browser allows users to explore large quantities of data, filtering the total number of results down into a more easily managed subset.

As we interact with an increasing number of domains, we find that the ease of use of the Semantic Forms throughout our sites greatly increases the quality of data that our users provide. Many of our projects start with relatively unstructured data, and after working with KEF, users have an easily managed and searchable repository of data.

We are continuing to evolve and mature the technology described in this paper. We already anticipate that work with evolving and new forms of social media, visual analytic tools, mobile devices, and additional collaborative tools (e.g., Drupal [57]) will continue to play an important role in our current and future projects.

Abbreviations

ADM: Automated Discovery Mechanism; BACH: Bayesian Analysis of Competing Hypotheses; BioCat: National Biosurveillance Integration System; CBR: Chemical Biological and Radiological; CMS: Content Management System; CPSE: Collaborative Problem Solving Environments; CSCW: Computer Supported Cooperative Work; DHS: Department of Homeland Security; DOE: Department Of Energy; EERE: Department of Energy's Office of Energy Efficiency and Renewable Energy; EPRI: Electric Power Research Institute; IED: Improvised Explosive Devices; ITD: Illicit Trafficking Demonstration; KEF: Knowledge Encapsulation Framework; MHK: Marine HydroKinetic; MIT: Massachusetts Institute of Technology; NER: Named Entity Recognizer; NLP: Natural Language Processing; PHPBB: PHP Bulletin Board; PNNL: Pacific Northwest National Laboratory; RSS: Really Simple Syndication; SME: Subject Matter Expert; SMW: Semantic Media Wiki; SRS: Scalable Reasoning System; TPA: Technosocial Predictive Analytics; TPAI: Technosocial Predictive Analytics Initiative; UNCC: University of North Carolina Charlotte; WYSIWYG: What You See Is What You Get.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

MM did provided significant contribution throughout the journal article and drafted the manuscript. AC and KF did much of the underlying research, and provided the information on the KEF process. RB provided the introduction, and information on other domain application. AP and PE provided use case and background information throughout. LM provided underlying research on NLP and provided information for the KEF model section. All authors read and approved the final manuscript.

Acknowledgements

This work was supported in part by the Pacific Northwest National Laboratory (PNNL) Technosocial Predictive Analytics Initiative. PNNL is operated by Battelle for the U.S. Department of Energy under Contract DE-AC06-76RL01830. The authors are indebted to reviewers and editors that have helped refine this paper and the associated research. PNNL Information Release No. PNWD-SA-9613.

Author details

[1]Pacific Northwest National Laboratory, 902 Battelle Boulevard, 999, MSIN K7-28 Richland, WA 99352, USA. [2]State of Washington, 735B Desoto Ave, Tumwater, WA 98512, USA.

Received: 11 October 2011 Accepted: 25 May 2012
Published: 22 August 2012

References

1. F Barjak, Research productivity in the internet era. Scientometrics **68**, 343–360 (2006)
2. KM Oliver, GL Wilkinson, LT Bennett, Evaluating the quality of internet information sources, in *ED-MEDIA & ED-TELECOM 97, June 14-19, 1997* (Association for the Advancement of Computing in Education, Calgary, 1997). http://www.eric.ed.gov/PDFS/ED412927.pdf
3. N Jennings, Twitter volume on the most tumultuous day of the campaign. Washington Post Blog (2012). 1/21/2012. Washington DC. Web. 3/12/2012. http://www.washingtonpost.com/blogs/election-2012/post/twitter-volume-on-the-most-tumultuous-day-of-the-campaign–atmentionmachine/2012/01/20/glQAkHVEGQ_blog.html
4. Twitter, *Twitter Numbers* (Twitter Blog, San Francisco, 2011). Web. 3/12/2012. http://blog.twitter.com/2011/03/numbers.html
5. A Sanfilippo, *Technosocial Predictive Analytics Initiative* (Pacific Northwest National Laboratory, Richland, 2011). Web.3/12/2012. http://predictiveanalytics.pnnl.gov
6. A Sanfilippo, AJ Cowell, L Malone, R Riensche, J Thomas, S Unwin, P Whitney, PC Wong, Technosocial predictive analytics in support of naturalistic decision making, in *9th Bi-annual international conference on naturalistic decision making (NDM9)* (BCS, London, 2009)
7. MC Madison, AK Fligg, AW Piatt, AJ Cowell, *Knowledge Encapsulation Framework* (Pacific Northwest National Laboratory, Richland, 2011). Web. 3/12/2012. http://kef.pnnl.gov
8. JP Ignizio, *Introduction to expert systems: The development and implementation of rule-based expert systems* (McGraw Hill, New York, 1991)
9. P Jackson, *Introduction to expert systems* (Addison Wesley, Boston, 1998)
10. AJ Cowell, ML Gregory, EJ Marshall, LR McGrath, Knowledge encapsulation framework for collaborative social modeling, in *Association for the advancement of artificial intelligence (AAAI)* (AAAI Press, Chicago, 2009)
11. AJ Cowell, KM Stanney, Manipulation of non verbal interaction style and demographic embodiment to increase anthropomorphic computer character credibility. Int J Hum Comput Stud Spec Issue: Subtle Expressivity for Characters and Robots **62**(2), 281–306 (2005)
12. EH Shortliffe, *Computer-based medical consultations MYCIN* (Elsevier, New York, 1976)
13. H Pople, *CADUCEUS An experimental expert system for medical diagnosis* (MIT Press, Cambridge MA, 1984)

14. R Sanguesa, J Pujol, Netexpert: Agent-based expertise location by means of social and knowledge networks, in *Knowledge management and organizational memories*, ed. by R. Dieng-Kuntz, N. Matta, First Editionth edn. (Springer, New York, 2002), pp. 159–168
15. P Dean, T Hoverd, D Howlett, *KnowledgeBench* (Cambridgeshire, United Kingdom,). Web. 3/12/2012. http://www.knowledgebench.com
16. RG Smith, JD Baker, The dipmeter advisor system: a case study in commercial expert system development, in *Proceedings of the eighth international joint conference on artificial intelligence (IJCAI'83), August 8-12, 1983; Karlsruhe, Germany* (Morgan Kaufmann Publishers Inc, San Francisco, 1983)
17. D Gracio, *Knowledge foundations & collaboratories: Bringing together people, tools, and science* (Pacific Northwest National Laboratory, Richland, 2008). http://www.pnl.gov/science/highlights/highlight.asp?id=225
18. RT Kouzes, JD Myers, WA Wulf, Collaboratories: doing science on the Internet. Computer **29**(8), 40–46 (1996)
19. I Gorton, C Sivaramakrishnan, G Black, S White, S Purohit, C Lansing, M Madison, K Schuchardt, Y Liu, A Velo, Knowledge-Management Framework for Modeling and Simulation. Computing Sci Eng **14**(2), 12–23 (2012)
20. VR Watson, Supporting scientific analysis within collaborative problem solving environments, in *HICSS 34 Minitrack on Collaborative Problem Solving Environments, January 3-6, 2001* (IEEE, Maui, 2001)
21. CA Shaffer, *Collaborative problem solving environments* (Virginia Tech, Blacksburg, 2008). Web. 3/12/2012. http://people.cs.vt.edu/~shaffer/Papers/DICPMShaffer.html
22. D Vesset, HD Morris, *The business value of predictive analytics* (IBM SPSS, San Jose, California, 2011)
23. SAS Institute Inc, *SAS architecture for business analytics* (SAS Institute Inc, Cary, NC, 2010)
24. A Vance, B Stone, Palantir the War on Terror's Secret Weapon, in *Business Week* (2011). Web. 3/12/2012. http://www.businessweek.com/magazine/palantir-the-vanguard-of-cyberterror-security-11222011.html
25. A Kittur, RE Kraut, Harnessing the wisdom of crowds in Wikipedia: quality through coordination, in *CSCW '08 Proceedings of the 2008 ACM conference on Computer supported cooperative work ACM, November 8-12, 2008; San Diego* (ACM, New York, 2008)
26. NS Friedland, PG Allen, G Matthews, M Witbrock, D Baxter, J Curtis, B Shepard, P Miraglia, J Angele, S Staab, E Moench, H Oppermann, D Wenke, D Israel, V Chaudhri, B Porter, K Barker, J Fan, SY Chaw, P Yeh, D Tecuci, P Clark, Project Halo: towards a digital Aristotle. AI Mag. **25**(4), 29–48 (2004)
27. S Singh, J Allanach, H Tu, K Pattipati, P Willett, Stochastic modeling of a terrorist event via the ASAM system, in *2004 IEEE International Conference on Systems Man and Cybernetics, October 10-13, 2004; The Hague, The Netherlands* (IEEE, Piscataway, 2004)
28. WikiMedia Project, Welcome to MediaWiki.org. (2012). Web. 3/12/2012. http://www.mediawiki.org
29. M Krötzsch, D Vrandečić, M Völkel, Semantic MediaWiki. Lecture Notes in Computer Science **4273**, 935–942 (2006)
30. Y Koren, *Semantic Forms* (2012). Web. 3/12/2012. http://www.mediawiki.org/wiki/Extension:Semantic_Forms
31. M Stefaner, User interface design, in *Dynamic taxonomies and faceted search: Theory,practice, and experience*, ed. by G. Sacco, Y. Tzitzikas. The Information Retrieval Series, Vol. 25 (Springer, New York, 2009)
32. Apache, *Apache SOLR*. Web. 3/12/2012. http://lucene.apache.org/solr/
33. D Pean, A Wright, J Phoenix, *Social Profile*, 2012. Web. 3/12/2012. http://www.mediawiki.org/wiki/Extension:SocialProfile
34. Ontoprise GmBH, *HaloACL* (, 2012). Web. 3/12/2012. http://www.mediawiki.org/wiki/Extension:Access_Control_List
35. *PHPBB*. (2012). Web. 3/12/2012. http://www.phpbb.com/
36. *Wordpress*. Web. 3/12/2012. http://wordpress.com/
37. B Carpenter, Phrasal queries with LingPipe and Lucene, in *Proceedings of the 13th Meeting of the Text Retrieval Conference (TREC), November 16-19, 2004; Gaithersburg, Maryland* (National Institute of Standards and Technology, Gaithersburg, 2004)
38. Amazon.com, Inc, *Amazon.com statistically improbable phrases*. Web. 3/12/2012. http://www.amazon.com/gp/search-inside/sipshelp.html
39. ML Gregory, LR McGrath, EB Bell, K O'Hara, K Domico, Domain independent knowledge base population from structured and unstructured data sources, in *Association for the Advancement of Artificial Intelligence* (AAAI Press, San Francisco, 2011)
40. A Sanfilippo, LR Franklin, S Tratz, GR Danielson, N Mileson, R Riensche, L McGrath, Automating frame analysis in social computing behavioral modeling and prediction, in *Social Computing, Behavioral Modeling, and Prediction*, ed. by H. Liu, J.J. Salerno, M.J. Young (Springer, New York, 2008), pp. 239–248
41. A Sanfilippo, B Baddeley, C Posse, P Whitney, A layered dempster-shafer approach to scenario construction and analysis intelligence and security informatics, in *IEEE International Conference on Intelligence and Security Informatics 2007 (ISI 2007), May 23-24, 2007; New Brunswick, NJ* (IEEE, Piscataway, NJ, 2007), pp. 95–102
42. Y Li, D McLean, ZA Bandar, JD O'Shea, K Crockett, Sentence similarity based on semantic nets and corpus statistics. IEEE Trans on Knowledge and Data Engineering **18**(8), 1138–1150 (2006)
43. T Pedersen, S Patwardham, J Michelizzi, WordNet: similarity measuring the relatedness of concepts, in *Proceedings of the nineteenth national conference on artificial intelligence (AAAI-04), July 25-29, 2004; San Jose, CA* (AAAI Press, Menlo Park, CA, 2004), pp. 1024–1025
44. V Kolhatkar, *An extended analysis of a method of all words sense disambiguation. MSc thesis* (Department of Coputer Science, University of Minnesota)
45. AJ Cowell, RS Jensen, ML Gregory, PC Ellis, K Fligg, LR McGrath, OH Kelly, E Bell, Collaborative knowledge discovery & marshalling for intelligence & security applications, in *2010 IEEE international conference on intelligence and security informatics (ISI) May 23-26, 2010; Vancouver BC Canada* (IEEE, Piscataway, NJ, 2010), pp. 233–238
46. R Riensche, LR Franklin, PR Paulson, AJ Brothers, D Niesen, LM Martucci, RS Butner, G Danielson, Development of a model-driven analytical game: Observations and lessons learned, in *The 3rd international conference on human centric computing (HumanCom 10) August 11-13, 2010; Cebu, Philippines* (IEEE, Piscataway, NJ, 2010)
47. R Riensche, PR Paulson, G Danielson, S Unwin, RS Butner, S Miller, LR Franklin, N Zuljevic, Serious gaming for predictive analytics, in *AAAI spring symposium on technosocial predictive analytics, March 23-25, 2010* (AAAI Press, Stanford, CA, 2009)
48. R Riensche, LM Martucci, JC Scholts, MA Whiting, Application and evaluation of analytic gaming, in *International conference on computational science and engineering (2009 CSE '09), August 29-31, 2009; Vancouver BC, Canada* (IEEE, Piscataway, NJ, 2009), pp. 1169–1173
49. CD Corley, RT Brigantic, M Lancaster, J Chung, C Noonan, J Schweighardt, S Brown, AJ Cowell, AK Fligg, AW Piatt et al., BioCat: Operational biosurveillance model evaluations and catalog, in *Supercomputing 2011 Computational Biosurveillance Workshop, November 12-18, 2011* (IEEE, Seattle, 2011)
50. R Anderson, A Copping, F Can Cleave, S Unwin, E Hamilton, *Conceptual model of offshore wind environmental risk evaluation system: Environmental effects of offshore wind energy fiscal year 2010. PNNL-19500* (Pacific Northwest National Laboratory, Richland, WA, 2010)
51. Pacific Northwest National Laboratory, *Signature Discovery Initiative* (Pacific Northwest National Laboratory, Richland WA, 2012). Web. 3/12/2012. http://signatures.pnnl.gov/
52. Pacific Northwest National Laboratory, *Carbon Sequestration Initiative* (Pacific Northwest National Laboratory, Richland WA, 2011). Web.3/12/2012. http://csi.pnnl.gov
53. Pacific Northwest National Laboratory, *Microbes FSFA* (Pacific Northwest National Laboratory, Richland WA, 2011). Web.3/12/2012. http://microbes.pnl.gov/wiki/
54. A Sanfilippo, L McGrath, P Whitney, Violent frames in action. Dynamics of Asymmetric Conflict: Pathways Toward Terrorism and Genocide **4**(2), 103–112 (2011)
55. W Pike, J Bruce, B Baddeley, D Best, L Franklin, R May, D Rice, R Riensche, K Younkin, The scalable reasoning system: lightweight visualization for distributed analytics. Inf Vis **8**(1), 71–84 (2009)
56. Pacific Northwest National Laboratory, *IN-SPIRE™ Visual Document Analysis* (Pacific Northwest National Laboratory, Richland WA, 2012). Web. 3/12/2012. [http://in-spire.pnnl.gov/]
57. Drupal Association, *Drupal*. Web. 3/12/2012. http://drupal.org

doi:10.1186/2190-8532-1-10
Cite this article as: Madison *et al.*: **Knowledge encapsulation framework for technosocial predictive modeling.** *Security Informatics* 2012 **1**:10.