

Automated Semantic Knowledge (ASK) Technology Overview

Success in business means staying ahead of the curve. It requires insight that will impact customers. Following the same path will never get you there. Only innovation will lead to the top.

The Right Question

The critical step to developing an innovative search infrastructure is asking the right question.

It's not about counting words.

It's not about how often words occur together.

It's not about words as nouns.

It's about understanding ideas and the relationships between them.

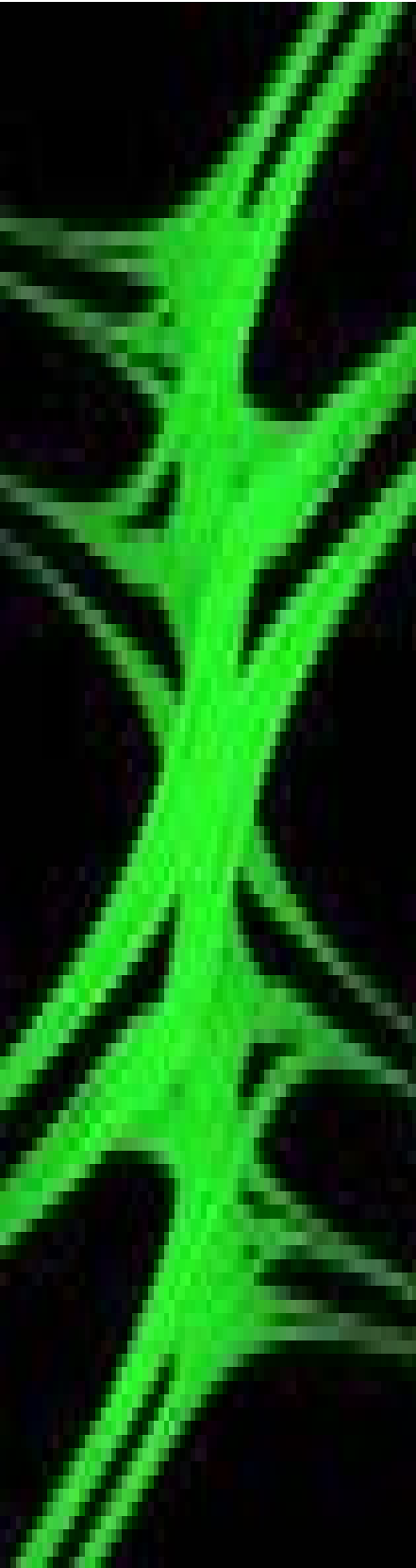
2028's Automated Semantic Knowledge (ASK) provides a better relevance metric and a substrate for great improvements to applications such as targeted advertising, directory augmentation, and automated filing.

2028's Automated Semantic Knowledge

ASK has taken three years to develop and addresses the heart of the information organization problem: How to make a system understand the essence of a document. Understanding is more than a collection of facts; rather, it is the relationships between them.

Designed to mimic human reading and analysis, ASK understands a document's semantic relevance by identifying and analyzing the concepts and their relationships via innovative data representations, set algorithms, and heuristics. Just as the Page Rank (Link Analysis) algorithm takes inspiration from the human heuristic of leveraging annotated references (links), ASK uses comparable insightful heuristics to dramatically improve intra-document analysis and search.

ASK amalgamates the individual pieces of information over a set of documents into a contextual lattice of knowledge. This powerful new substrate provides a platform for intelligently processing unstructured information, giving it the ability to address the needs of a variety of different applications.



ASK technology provides:

Accuracy. ASK's humanistic relevance metric provides a more intuitive ranking. This is especially true given instances where additional information (such as links) is not available. In contrast, if such information is available ASK can be synergistically combined with other algorithms such as Page Rank.

Speed and Scalability. With the volume of data today, speed and scalability are just as important as accuracy. Unlike many other solutions, ASK does not require manual tagging or example pre-classified document sets, enabling vast amounts of data to be processed in a short period of time. Also, ASK leverages standard off-the-shelf SQL Database solutions as a backend, and thus can provide solutions for terabyte-sized data. To facilitate rapid scaling ASK can be easily run in parallel.

Easy Integration. ASK was designed to work in collaboration with other applications. Accordingly, it provides both XML and RDF (Resource Description Framework) interfaces as defined by the W3C to ease the integration within the workplace. Through standard interfaces, ASK allows other workplace applications over the Internet or Intranet to

Example Applications

- Security - Identify unusual content access behavior.
- Collection Intelligence- analyze sets of documents to get a high level view
- Content Routing - get the right content to the right people
- Content Profiling - thoroughly understand a document and its relationship to the big picture
- Information Portals - quickly design and deploy targeted portals
- Knowledge Management - focus on learning and innovating - not organizing
- Online Publishing - automate the organization of published content, revolutionizing the way users interact with content
- User Profiling - understand user behavior at a whole new level
- User Content Interaction - analyze and understand how users and content interact

Given a document or set of documents, ASK can:

- Identify the concepts and their significance to the documents
- Determine how concepts are related
- Create automated adaptive concept maps
- Derive relations between documents
- Determine unique pages
- Spider dynamic web sites
- Automatically generate a directory structure
- Use embedded information in documents, such as links, to further refine understanding

leverage its structured view of textual information in a straightforward and logical manner.

Real-time categorization. The semi-structured information generated by ASK can be used to generate adaptive directories that reflect the content available. This real-time categorization enables fully automated organization of information that continually adapts to both the content and user.

Content formats. ASK supports a variety of content formats: HTML, PDF, PS, and DOC.

In summary, ASK alleviates the burden of content organization by automating tasks previously needing human effort. The derived structured information (concepts, concept relations, dates, etc.) can then be used to generate key concepts, synopses, automated indexes for books, and retrieve semantically similar pages.

ASK provides the infrastructure necessary to automate the routing, organizing, and retrieving of unstructured information in an organized, efficient, and meaningful manner. Through combining this data, ASK automatically generates new knowledge, such as concept maps and provide intelligence tools to help analyze trends, user interests and other information allowing you to understand your content, your users, and how they work together.