



Benefits

- **Reliable results:**
Delivers more relevant search results, with fewer queries by the user, compared to other search engine technologies
- **Contextual relevance:**
Enables users to discover words, ideas, and situational details that are contextually associated with a specific query
- **Intelligent search:**
Allows users to discover key themes in large document sets, with no prior knowledge of the documents
- **Efficient classification:**
Eliminates the time and expense associated with maintaining document categorization (i.e., ontology), delivering semantic and conceptual search results

Technology Details

Perilog's underlying algorithm is based on the theory of experiential iconicity, which states that patterns of relatedness among things in the world of experience systematically influence patterns of relatedness among words in written discourse. Perilog's ability to deliver semantic and conceptual search results through an automated algorithm, without the need to rely on natural language processing or manually (or semi-manually) maintained categorization, follows directly from the theory of iconicity.

How it works

Perilog measures the degree of contextual association of large numbers of term pairs in text to produce network models that capture the structure of the text and, by virtue of Perilog's validated theory of iconicity, the structure of the domain(s), situation(s), and concern(s) expressed by the author(s) of the text. In fact, given alphanumeric representations of any other sequences in which context is meaningful—such as music or genetic sequences—Perilog can derive their contextual structure.

Operating on a document set (i.e., corpus) or a single document, Perilog creates a network model of contextually related words and phrases. When a user enters a keyword or key phrase search, Perilog creates a query network of “topical hubs,” based on the query words input by the user. Phrases may be of any number and length. Each phrase is represented by a network, and these networks are combined into a single query network.

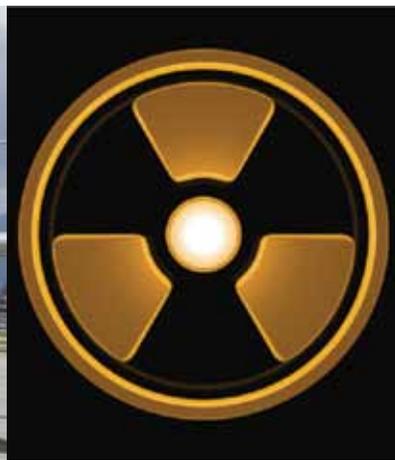
By matching the phrase query network with document networks, Perilog's phrase search provides flexible and thorough phrase matching that is unavailable with other methods. Instead of the keyword search being limited to the query words alone, Perilog uses the relationships of keywords within their contextual associations to find documents in which those relationships are significant.

Key features and methods

Perilog's key features and methods encompass text analysis, modeling, relevance-ranking, keyword and phrase search, phrase generation, and phrase discovery:

Text Analysis: The process converts bodies of text to sequences of terms and measures the contextual associations among them. This determines the structure of text as a way of measuring the structure of the domains and situations represented by the text. Terms that are contextually related in the structure of text are considered contextually related in the world represented by the text.

Modeling: Each Perilog model consists of a network of contextually associated terms. A Perilog model can represent any body of text, from an entire database to a short phrase, and



it can represent any domain, sub-domain, situation, situational detail topic, or subtopic.

Relevance Ranking: Perilog quantifies the similarity of any two models by comparing their paired terms and contextual measurements. One model's features are used as relevance-ranking criteria and compared to a collection of models, enabling the models in the collection to be ranked according to their relevance to the criteria. By ranking a collection of models on every model in the collection, an association matrix can be created to provide data for input to clustering methods.

Keyword and Phrase Search: Perilog retrieves from a user-specified database documents that contain one or more user-specified keywords or phrases in typical or selected contexts, and ranks the documents on their relevance to the keywords or phrases in context. The most relevant documents are automatically highlighted and displayed in a Web browser window, allowing the user to scroll through and review them.

Each of the documents is accompanied by a list of related words or phrases that contribute to the relevance of the document. Experienced users can refer to these relations to understand which features were interpreted as contributing to the relevance of the document. In some cases, this can lead the user to modify or fine-tune the search strategy.

Phrase Generation: To aid a search, Perilog can produce a list of phrases from the database that contain a user-specified word

or phrase that can be used to suggest queries for phrase searches. To generate phrases, the user provides a word or phrase that is to be contained in each of the output phrases. Perilog builds phrases around this input, based on its phrase models. The resulting phrases are displayed on the computer screen or can be redirected to a file. The phrases are sorted based on an estimate of their prominence in the document set.

Phrase Discovery: Further aiding the relevance of search results, Perilog can find phrases that are related to topics of interest. For example, given a topic such as "fatigue," Perilog can discover related phrases such as "rest period," "reduced rest," "duty period," and "crew scheduling." Phrase discovery can help users understand the variations and scope of topics in a document set. They also can be used selectively as input to a Perilog phrase search, enabling retrieval of documents that contain particular topical variations.

The first step in phrase discovery is to perform a keyword or phrase search. Next, phrases are automatically extracted from the most relevant documents. The phrases produced at this point may be useful, but further processing will improve the results. From these phrases, topical phrases are distilled by a combination of manual and automated methods. The refined set of topical phrases then can be used to query the database, using phrase search. The cycle of phrase extraction and search is repeated to produce a final set of documents. If documents relevant to the topic are available

Applications

Perilog can enhance many search-related applications, including:

- Large knowledge management and document retrieval systems, for legal research, market research, intellectual property asset management, claims management, etc.
- Life sciences and medical research
- Intelligence analyses
- Commercial search engines
- Contextual online advertising
- Airline flight safety databases

in the document set, this final collection of documents will be highly relevant to the topic.

The main product is a list of topical phrases that are extracted from the final collection of documents.

Why it is better

Perilog overcomes many of the fundamental flaws that hinder more basic search engines:

- Reliance on the “bag of words” model, in which all words are treated equally, and relevance determined only by the frequency with which a keyword or phrase appears in the text
- Word ambiguity, or the inability to distinguish between different meanings of the same word (e.g., “bark”), producing ambiguous results for such queries
- Term mismatch, when the search engine yields only a fraction of relevant results because users select the same term to describe an object less than 20 percent of the time
- Query drift, when automatic expansion of queries results in unexpected and incorrect results

Perilog helps improve search results by addressing each of these flaws, determining the contextual association of words and word pairs in documents.

This network model enables Perilog to answer questions such as:

- What are the most prominent topics in this document (even if the user has never seen the document before)?
- What are the most relevant sections of the document regarding a particular topic?
- What other topics are related most closely to this topic?

Extensive manual intervention would be required to answer these questions using traditional search engines, as well as possible repeat searching and manual review of potentially

relevant documents. In contrast, Perilog is extremely useful for research-intensive, open-ended queries because of its capability to make contextual associations across a network model of relevant text, and requiring no prior knowledge of the document contents. Likewise, Perilog is very useful when searching domains where terminology changes frequently—in such areas as life sciences, medicine, and intelligence. The ability to intuit contextual associations enables search engines to deliver on the promise of conceptual search and semantic search.

Perilog’s keyword and search retrieval capabilities are augmented by its keyword and phrase discovery features, which can be used to discover and extract keywords and key phrases from an existing document or document set. This enables examination of a source document with other contextually relevant target documents in a document set. This feature is similar, on the surface, to the “Find More Documents Like This” feature in other search engines, but without Perilog’s phrase discovery method, other search engines are still subject to the flaws described above.

In short, Perilog facilitates a move from traditional keyword retrieval search to full conceptual and semantic searching, yielding more complete and relevant results in less time and with less manual intervention.

Patents

NASA has secured four patents that comprise the Perilog software: U.S. Patent Nos. 6,823,333; 6,741,981; 6,697,793; and 6,721,728.

Licensing and Partnering Opportunities

This technology is part of NASA’s Innovative Partnerships Program, which seeks to transfer technology into and out of NASA to benefit the space program and U.S. industry. NASA invites companies to inquire about the licensing possibilities for the Perilog technology (ARC-14512-1, ARC-14513-1, ARC-14514-1, and ARC-14515-1) for commercial applications.

For More Information

If you would like more information about this technology, please contact:

Entrepreneurial Initiatives Division
NASA Ames Research Center
technology@mail.arc.nasa.gov