

InnovationQ Plus Search Engine Technology

Introduction

The InnovationQ Plus search engine is powered by IP.com's patented neural network machine learning technology. This state-of-the-art system allows it to match queries to documents based on *meaning* rather than keywords. The result is a search that delivers a more complete result set as well as one with less noise or false positives. In technical information retrieval terms, this means it delivers superior *precision and recall*. It is able to do this because meaning or concept matching overcomes the inherent ambiguities in ordinary language, especially synonymy and polysemy. InnovationQ Plus understands that two terms—"vehicle" and "car"—have similar meanings (synonymy) while a single term like "stream" may have multiple unrelated meanings (polysemy). It uses this knowledge to retrieve the best, most relevant results to a query.

Machine Learning

Machine learning is a branch of artificial intelligence that employs statistical models and advanced algorithms to enable computers to become "intelligent" by "learning" from data. Rather than encoding a computer with a set of facts about the world and a set of logical reasoning rules for inference, machine learning systems detect significant patterns from the processing of large data sets. Although there are many types of machine learning techniques, those based on neural networks have recently gained prominence and are delivering impressive results in a range of applications including voice and image recognition, language translation, sentiment analysis and natural language processing (NLP). InnovationQ Plus applies neural network machine learning technology to understand the meaning of documents and queries.

Computational Meaning

InnovationQ Plus discovers meaning or semantics through the statistical analysis of word patterns and distributions natively occurring in massive collections of documents. External thesauri or ontologies aren't used. Instead, meaning is derived by analyzing (through the neural net machine learning system) the large scale probability and distributional properties of words in the InnovationQ Plus document collections. The derivation of meaning from these patterns is well supported in the technical literature and is referred to as the "distributional hypothesis of meaning":

The Distributional Hypothesis: words with similar distributional properties have similar meaningsⁱ

This statistical approach currently dominates the field of NLP and there is an enormous literature on it as well as rapidly advancing practical applications from the most advanced technology companies in the world including Google, Apple, Facebook and Amazon.

The InnovationQ Plus machine learning engine captures these statistical patterns and uses them to convert both documents and queries to mathematical objects known as feature vectors. The vectors encode statistical properties of words including co-occurrence pattern counts, frequency counts and word sequence probabilities. InnovationQ Plus also includes an innovative type of back-off or error correction to supplement or blend the semantic feature vectors with keyword representations when necessary. Documents and queries with similar vector scores are closely related semantically, regardless of the literal composition of the words they contain. In retrieval the system uses these vectors to compute the probability of a match between document and query.

Self-Learning

As a document collection grows over time, new terms and concepts inevitably emerge while others may shift in meaning. For example, phrases like “internet of things”, and “deep learning” have gained currency in the technical literature in recent years. The same is true for individual words like “multicore” or “qubit”. InnovationQ Plus automatically adjusts to these changes in the topical composition of the document collection. It detects this “semantic drift” through its learning algorithm and re-weights concepts and topics accordingly. This makes the system self evolving and allows it to powerfully and elegantly sustain high precision and recall in the retrieval task as the balance of topics and concepts in a document collection changes.

Semantic Search and Knowledge Representation

Semantic search is itself an ambiguous phrase. Some vendors use it to refer to a search that returns a knowledge representation based on entity and fact extraction. These systems rely on ontologies to define the meaning of terms and typically present graph based results using subject/predicate/object triples. While these systems deliver value beyond traditional keyword retrieval they typically require “structured” text that has been manually annotated with a machine readable industry standard mark-up language such as schema.org or the Resource Description Framework (RDF) specifications from the World Wide Web Consortium (W3C). Search engines can use these marked up pages to populate “knowledge panels” or “knowledge graphs” in response to certain types of queries.

InnovationQ Plus attacks a different problem in a different way. It doesn’t try to extract entities and facts. It doesn’t rely on machine readable semantic mark-up. It processes unstructured, unlabeled text collections and automatically distills the key concepts inherent in the statistical distribution of the words in the documents.

Key Questions

When evaluating systems claiming semantic capabilities there are some key questions to ask:

1. Is the entire document collection semantically searchable? In the case of InnovationQ Plus the answer is yes. Other vendors may only make a portion of their collection semantically searchable.
2. Does the vendor own its semantic technology? For InnovationQ Plus the answer is again yes. Other vendors may license their technology from a supplier and may not control the shape and pace of enhancements.
3. What technology does the vendor use to power semantic search? InnovationQ Plus uses the latest neural network machine learning technology. Other vendors may still be using Latent Semantic Indexing as the basis of a semantic search. This is an older technology developed in the late 1980s with known limitations.
4. What is the response time for semantic search? InnovationQ Plus presents results in 2-4 seconds or even faster depending on the network connection. Other vendors' semantic searches can take much longer.
5. Has semantic search been added to an older keyword system? InnovationQ Plus has been designed from the bottom up exclusively around semantic search. Other vendors have added semantic search to legacy systems often resulting in sub-optimal performance and usability.
6. How are queries processed? InnovationQ Plus applies the same neural network learning algorithms to queries and documents. The result is that InnovationQ Plus can handle a variety of long queries; paragraphs or entire pages of text can be entered as a query eliminating the need for the searcher to select key words or phrases. Other vendors may limit their semantic processing to documents alone.

**Learn More about
InnovationQ Plus**

Contact us:

- www.ieee.org/innovationqplus
- +1-800-701-4333 (USA/Canada) // +1-732-981-0060 (worldwide)
- onlinesupport@ieee.org

ⁱ Sahlgren, M. (2006). *The Word-space model* (Doctoral dissertation, Gothenburg University). p.21