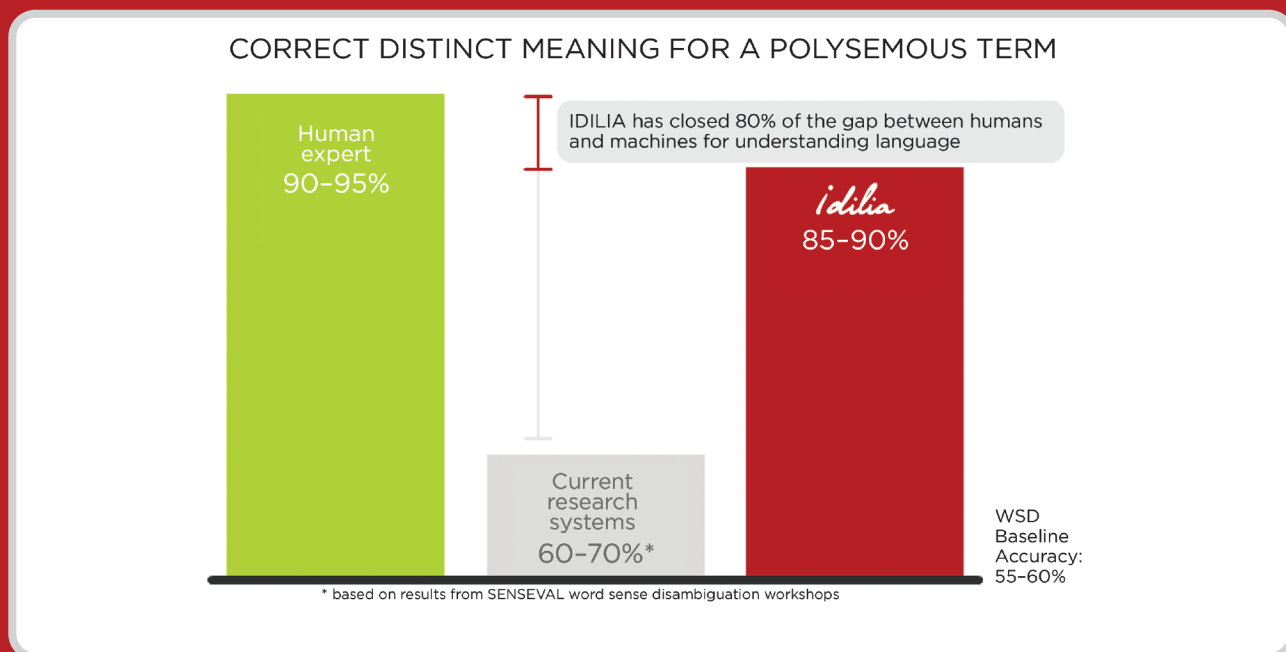


Idilia has achieved a scientific breakthrough in the field of artificial intelligence by developing technology that allows software to understand and use the precise meanings of words in naturally expressed language (e.g. to understand when “bank” refers to a river, versus when it means a financial institution). This process is known as **word sense disambiguation (WSD)**.

For the past fifty years, no technology has been able to achieve greater than 60-70% accuracy in word sense disambiguation<sup>1</sup>. This level of accuracy provides almost no additional value beyond selecting the most common word sense in every instance. Idilia’s WSD technology has decisively broken through this ceiling and is 85-90% accurate – within 5-10% of human performance. This means that Idilia has closed 80% of the gap between humans and machines for understanding language.

This is the world’s first system capable of performing WSD at accuracy approaching that of a human being and represents a milestone in the field of artificial intelligence.



Accurate word sense disambiguation enables software applications to correctly interpret the meaning of words based on their context. In application, this technology will dramatically improve the accuracy of Internet search, machine translation and automatic speech recognition, among other applications.

A few key features of the Idilia WSD system include:

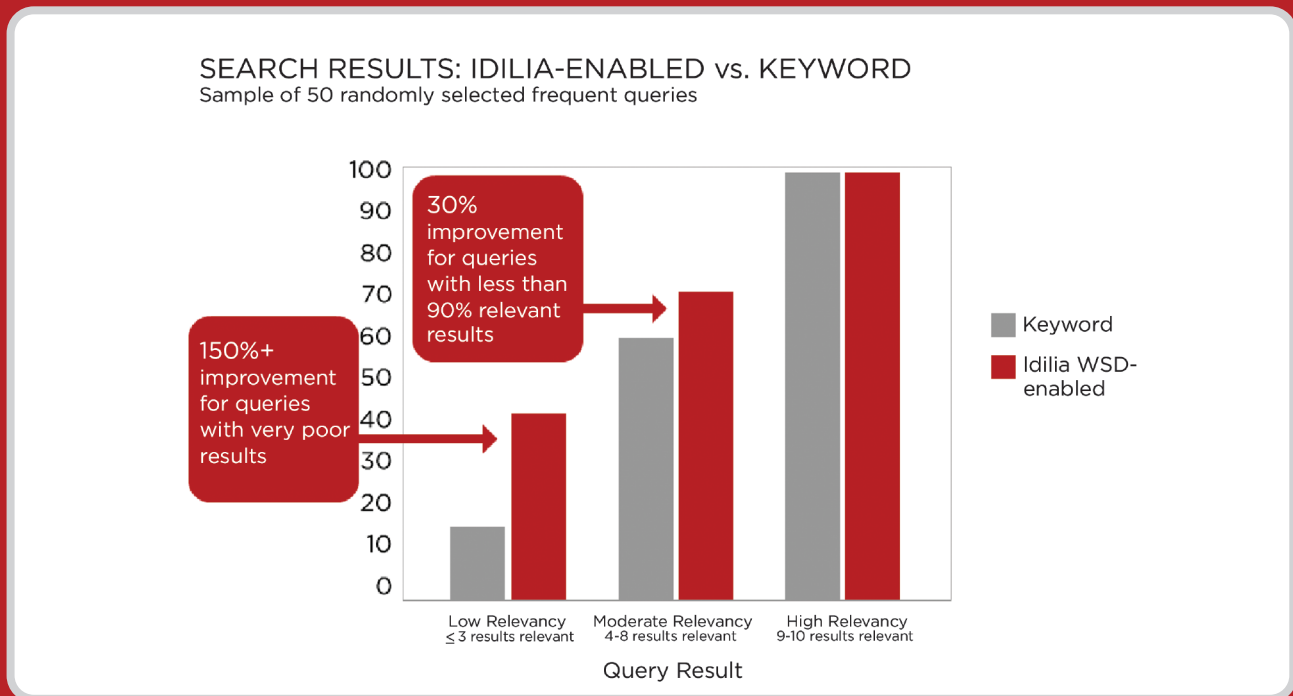
- Significant research breakthroughs in linguistic algorithms and word sense disambiguation methods.
- A first-of-its-kind dispute resolution algorithm that arbitrates between semantic results provided by multiple algorithms and linguistic components.
- The first WSD system to seamlessly integrate named entity recognition (capable of recognizing approximately 500 categories of named entities) allowing accurate interpretation of proper noun and adjective senses.
- The largest linguistic knowledge base ever created, containing over 20 million nodes (representing over 5 million proper nouns and 200,000 common word senses), connected by over 50 million edges (covering a range of over 1,000 types of precisely specified

<sup>1</sup> “At present, WSD work is at a crossroads: systems have hit a reported ceiling of 70%+ accuracy (Edmonds and Kilgarriff 2002)...” Eneko Agirre and Philip Edmonds, *Word Sense Disambiguation Algorithms and Applications*: Springer, 2006

semantic relations). By way of comparison, Wikipedia currently contains approximately 2 million articles covering roughly half the number of concepts.

- Over 5 million words of precisely sense-tagged training text, used to train the machine learning based algorithms. Developed using proprietary tools, this is over 5 times the amount of training data available to other systems.
- Automated system learning based on user feedback. In a deployment, the accuracy of the system will continuously improve in response to user feedback, conferring an ever-widening first-mover advantage.

Idilia's technology has significant and immediate commercial applications, most notably Internet search, internet advertising, and information retrieval. Idilia's WSD enables search engines to match the senses of words in a query precisely to the senses of words in a document or paid advertising listing. This functionality offers a fundamental improvement in relevancy over current search techniques. In tests utilizing a leading commercial search engine, Idilia's technology improved relevancy by more than 150% for queries for which the initial search engine results were very poor, and by 30% for queries for which the search engine results were of moderate relevancy. In addition, the technology is expected to improve on-topic first-page paid advertising listings by 20-25%.



Beyond search, WSD technology will facilitate deployment of speech recognition systems that perform like people. As well, WSD opens the multi-billion dollar translation market to viable software applications, because, by resolving meaning, it enables machine translation software to attain commercially acceptable levels of accuracy.

Idilia has been backed by The Hearst Corporation and the National Research Council of Canada, the Canadian federal government's research arm. Idilia was founded in April 2000 based on research begun in 1998. Idilia's team is comprised of the largest concentration of WSD talent in the world, including machine learning and linguistics experts with over 300 years of combined commercial R&D experience. The company has invested over 150 person-years in WSD research and development and has four patents pending against critical aspects of the technology.