

# Computer scientists improve access to millions of US patents records

January 7 2016, by Janet Lathrop

---

Thanks to the work of computer science researchers at the University of Massachusetts Amherst, the online database of the country's millions of inventors and patents will be much better to navigate for innovators, business and policy makers.

A team of UMass Amherst computer scientists took the top prize in an international competition sponsored by the U.S. Patent and Trademark Office (USPTO) and U.S. Department of Commerce. They designed a computer algorithm that rapidly removes inventor ambiguity from patent records, which will provide users more efficient and effective searches. Their winning approach will be incorporated into the USPTO's new online platform, PatentsView.

The team, advised by professor Andrew McCallum's Information Extraction and Synthesis Laboratory, produced the winning algorithm to take the top prize, which attracted entries from China, Germany, Australia, Belgium and the United States. McCallum, who is also the director of the UMass Center for Data Science, says other team members included graduate students Nicholas Monath and Ari Kobren; Michael Wick, now at Oracle Labs; Sameer Singh, now at the University of Washington, and Jack Sullivan, now at Cambridge Semantics.

Undersecretary of Commerce for Intellectual Property and USPTO director Michelle K. Lee said the goal of the workshop was "to encourage the development of novel approaches to reveal inventor identities across nearly 40 years of U.S. patent data." It will provide

users more efficient and effective searches of the country's millions of inventors and patents.

As part of its win, the lab will receive a \$25,000 stipend for technical guidance on applying the "entity disambiguation" algorithm to millions of patent records in the PatentsView platform.

"Winning was a great honor and it's a thrill to know this research will be used by the USPTO for the public good," says Monath, a native of Harvard, Mass. "The workshop also provided the opportunity to talk with people outside of computer science, in economics and policy, who rely on inventor data, which gave me new perspective for my research."

The UMass team took the top prize because its solution excelled in time and accuracy. "We had the fastest system in the competition and the system with the highest accuracy score," Monath says.

Entity disambiguation means differentiating among many individuals (entities) with similar attributes and grouping them together correctly, that is, unambiguously, says Monath, who studies machine learning and natural language processing. Inventor disambiguation is important to the USPTO because inventors often appear in patent records with different names, spellings and nicknames or because multiple inventors may have the same name.

Such ambiguities in the current system make data queries unreliable, requiring time-consuming manual intervention, the computer scientists explain. The competition asked contestants to "disambiguate" the inventors of over 12 million patent records filed between 1976 and 2014. Given the large size of the data set, manual data resolution would be unreasonable, which is why automated methods are necessary, they add.

Monath says, "Our method uses a hierarchical approach to

disambiguation, which has several advantages over alternative pairwise approaches. Our method considers groups of two or more mentions in determining the disambiguation and has a more efficient disambiguation procedure."

Russell Slifer, USPTO deputy director, said at the competition workshop that PatentsView, a new web tool for the public to explore its warehouse of patenting data in the U.S., is intended to help innovators, business and [policy makers](#) by democratizing USPTO data. The office had always made 225 years of patent and trademark data available to the public, but it was difficult to use because of inconsistent formatting and complicated data formats. These cause inventor ambiguities, making the data useable only by the few people willing to spend the required time and effort.

PatentsView provides a more accessible online interface for patent searches. It allows any user to explore technological, regional and individual-level patent trends via search filters with multiple viewing options.

**More information:** [www.dev.patentsview.org/workshop/index.html](http://www.dev.patentsview.org/workshop/index.html)

Provided by University of Massachusetts Amherst

Citation: Computer scientists improve access to millions of US patents records (2016, January 7) retrieved 11 May 2024 from <https://phys.org/news/2016-01-scientists-access-millions-patents.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.