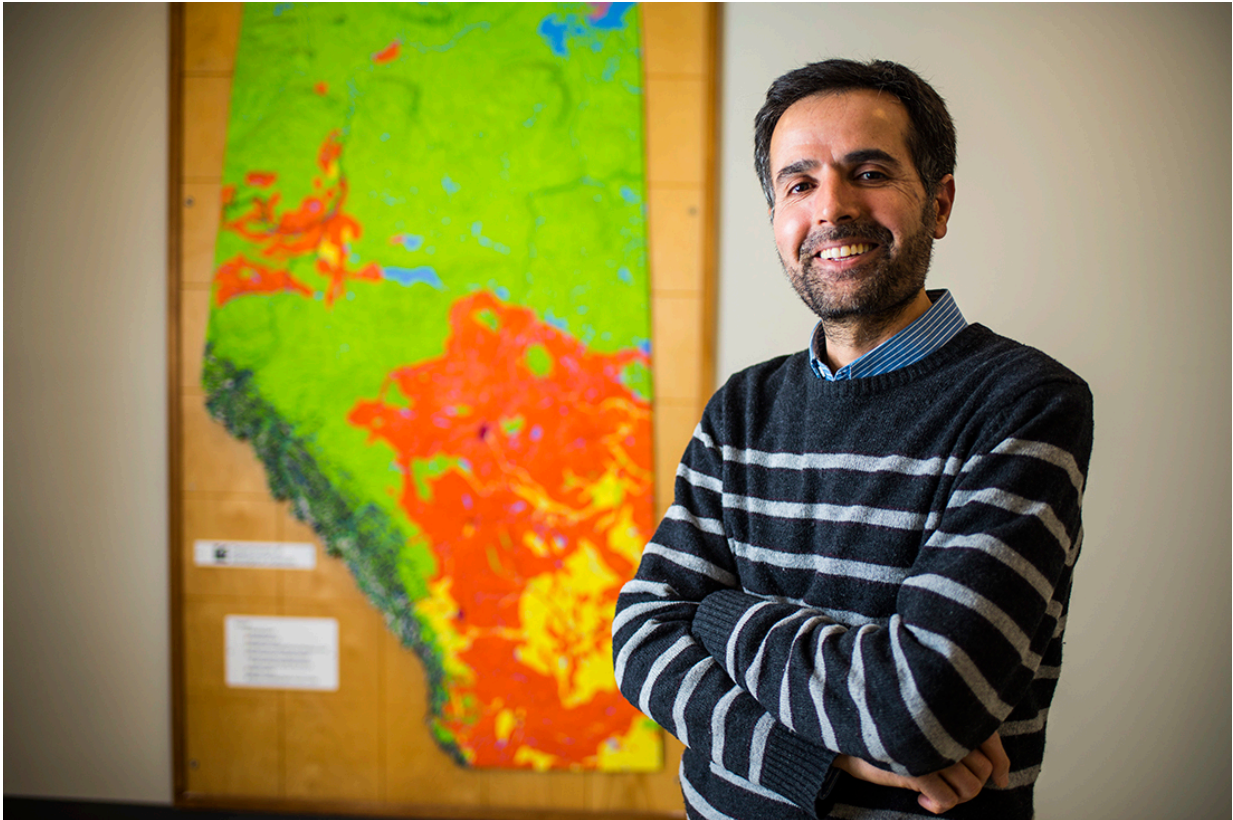# How on earth does geotagging work?

January 9 2017, by Katie Willis

Davood Rafiei is a professor in the Department of Computing Sciences and expert in big data and information management. Credit: John Ulan

In an increasingly digital world, we don't always consider where on earth the information we find online comes from.

Now, computing science researchers at the University of Alberta are

using automated geotagging models to put a place to online data and documents.

"With the proliferation of online content and the need for sharing it across the globe, it is important to correctly match names to the places they refer to," says Davood Rafiei, professor in the Department of Computing Sciences and expert in big data and information management.

"The potential applications are huge. Perhaps you want to find out about people, organizations, or events in a certain location. Or maybe you want to understand where your data sources are located. There are even applications for determining if two named entities are in fact referring to the same thing."

Using a two-part model, Rafiei and former master of science student Jiangwei Yu have developed a technique to automate geotagging for news articles and other online documents and data. The model integrates two competing hypotheses: inheritance and near-location.

According to the inheritance hypothesis, named entities are given the same geographical location as the document in which they are mentioned. "For example, every name mentioned in a Wall Street Journal article will inherit the geocentre of the article, which in this case will be New York City, New York, USA," explains Rafiei.

The near-location hypothesis links the named entities to geographical locations mentioned in nearby text—such as a person's name mentioned next to the phrase "Edmonton, Alberta" in an article.

"What happens in the real world though appears to be a mixture of the two forces," explains Rafiei. "Our data shows that the inheritance hypothesis holds in 72 percent of the cases, the near-location hypothesis

holds in 67 percent of the cases, and at least one holds in close to 99 percent of the cases."

In addition to being highly accurate, the model is automated, cutting the cost of geotagging significantly.

"The power of geotagging is being better able to understand people, places, and things referenced in online documents," says Rafiei.

The paper, "Geotagging named entities in news and online documents", was presented at International Conference on Information and Knowledge Management, Proceedings.

 More information: Geotagging named entities in news and online documents, DOI: 10.1145/2983323.2983795

Provided by University of Alberta