Guided Discovery of Information

Technology #2018-053

Current information discovery functions involve use of keyword searches that return results matching one or more terms that a user specifies. A drawback with this approach is that the results returned depend on the user’s selection of terms versus identification of documents that are of greatest interest. This means that if a user does not know that a particular term will return a desirable result, the user may never receive the information they seek.

Invention Summary:

Researchers in at Rutgers University have invented an information discovery technique that optimizes the chances that a user will receive a document that is closest to a concept of interest. The goal is to guide people’s exploration of a domain such as text. When provided with an initial datum, such as a word describing a general concept of interest, this technique suggests subsequent candidate data for a user to choose among. Suggestions are designed to converge rapidly on the source of greatest interest for the user, differentiating amongst the documents that are consistent with the previous elements of the query.

This invention operates by leveraging a new mathematical theory that formalizes optimal conditions for learning in cooperative settings with the goal of retrieving a specific source of information. The invention implements these theoretical ideas in an extended setting using communication that occurs between two agents, which we call a teacher and a learner. Here the teacher represents the process of selecting data to convey a concept of interest, and the learner represents the inference process of interpreting the received data. The invention describes human inference in cooperative situations. This is used in a novel and surprising way—to facilitate search through massive text corpora. The approach integrates breadth and depth first search by considering words whose semantics overlap but also have unique aspects.

Advantages:

• Facilitates searching for information in the form of a database, texts, collection of images or videos, the internet, any other source of quantified information (i.e. data) or a combination of these.

• Supports rapid and effective search through massive repositories of information.

Applications:

• Industries that involve information discover of massive amounts of data (finance, health, government, legal, IoT, etc.)
• Search algorithms such as information filtering (the internet), recommender systems (Amazon, Netflix), etc.

**Intellectual Property & Development Status:**

The technology is patent pending and is currently available for licensing.

**Inventors**

**Patrick Shafto, PhD**

Dr. Patrick Shafto is the Henry Rutgers Term Chair in Data Science and Associate Professor in the Department of Mathematics and Computer Science at Rutgers - Newark. He is also affiliated with the Institute for Data Science, Learning and Applications (I-DSLA) and has appointments in Psychology, Rutgers Business School, and the Center for Molecular and Behavioral Neuroscience (CMBN).