INSTITUTIONEN FÖR DATAVETENSKAP | LUNDS TEKNISKA HÖGSKOLA | PRESENTERAD 2019-08-30

EXAMENSARBETE AI & knowledge graphs: NLP and graph analytics to structure data and reveal insights
STUDENT Iván Llopis Beltrán
HANDLEDARE Pierre Nugues (LTH)
EXAMINATOR Elin Anna Topp (LTH)

Artificial intelligence for document modelling

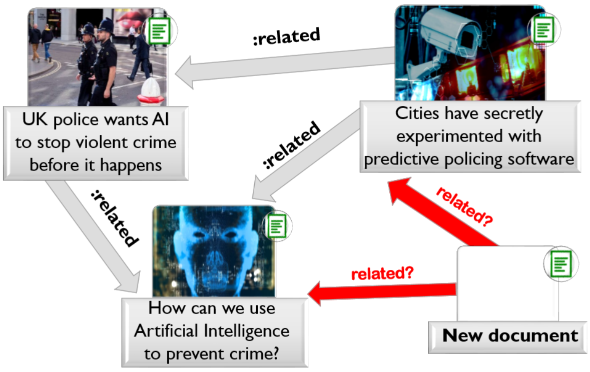
POPULÄRVETENSKAPLIG SAMMANFATTNING **Iván Llopis Beltrán**

Finding the information one wants is sometimes difficult. This project presents a way to characterize documents, extract useful information and find relations to other documents automatically.

One of the most common forms of communication is text. We interact with some sort of written form every day. Some times we need to go through tedious reads in order to look for something specific. Imagine you could get help with that: know what you want and help you find what you need. This project is about developing an intelligent system that is able to characterize documents by topics and automatically establish relations among them. Imagine you could go to a book store and someone could recommend you a book that talks about the topics that need or love so much.

This is achieved through artificial intelligence, or more specifically, topic modelling. Topics are the inherent themes that define what a document is about. We use this information to characterize documents and predict relations to other documents. For example, picture that you are an employer looking for the perfect candidate that matches an open position at your company. The algorithms described could help to sort CVs by similarity to the job description. It could also help a job seeker to find suitable offers that fit their profile more easily.

The system is also capable of extracting the keywords of texts. The analysis of such keywords summarize what is included in them and helps to understand the important information faster. The system is also able to find groups of documents that are interrelated among themselves. These so-called *communities* help obtain recommendations of related elements.


:related
:related
:related
UK police wants AI to stop violent crime before it happens
Cities have secretly experimented with predictive policing software
How can we use Artificial Intelligence to prevent crime?
related?
related?
New document

We analyze the commonalities of documents though three parameters: keywords, topics and context. The best result is achieved by comparison of specific words, rather by the other elements separately. However, close results are also achieved when looking at all three parameters combined.