



Amsterdam
Data Science

Can We Detect Changes in Concepts over Time from an Collection of Annotated Articles?

Oliver Becher, Laura Hollink & Desmond Elliott

16.05.2017

partners



UNIVERSITEIT VAN AMSTERDAM

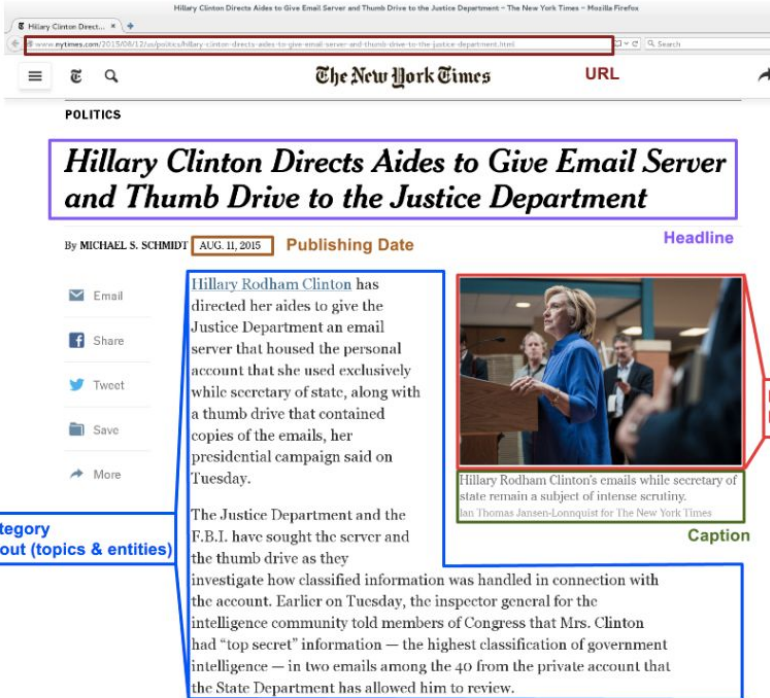


Hogeschool van Amsterdam

CWI

Data and Definition

The ION Dataset



The screenshot shows a New York Times article titled "Hillary Clinton Directs Aides to Give Email Server and Thumb Drive to the Justice Department". Annotations include:

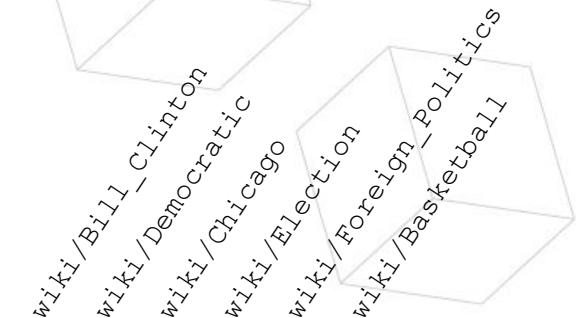
- Category About (topics & entities)**: A blue box pointing to the article title.
- Image URL** and **Image Features**: A red box pointing to the image of Hillary Clinton.
- Caption**: A green box pointing to the caption below the image.
- Text**: A blue box pointing to the main body text of the article.

Hollink; Elliott et al., A corpus of images and text in online news, 2016

What is a Concept?

Our practical definition:

- Each unique wikipedia annotation of a newspaper article
- Representation: Vector of co-occurrences with other concepts
- e.g.:



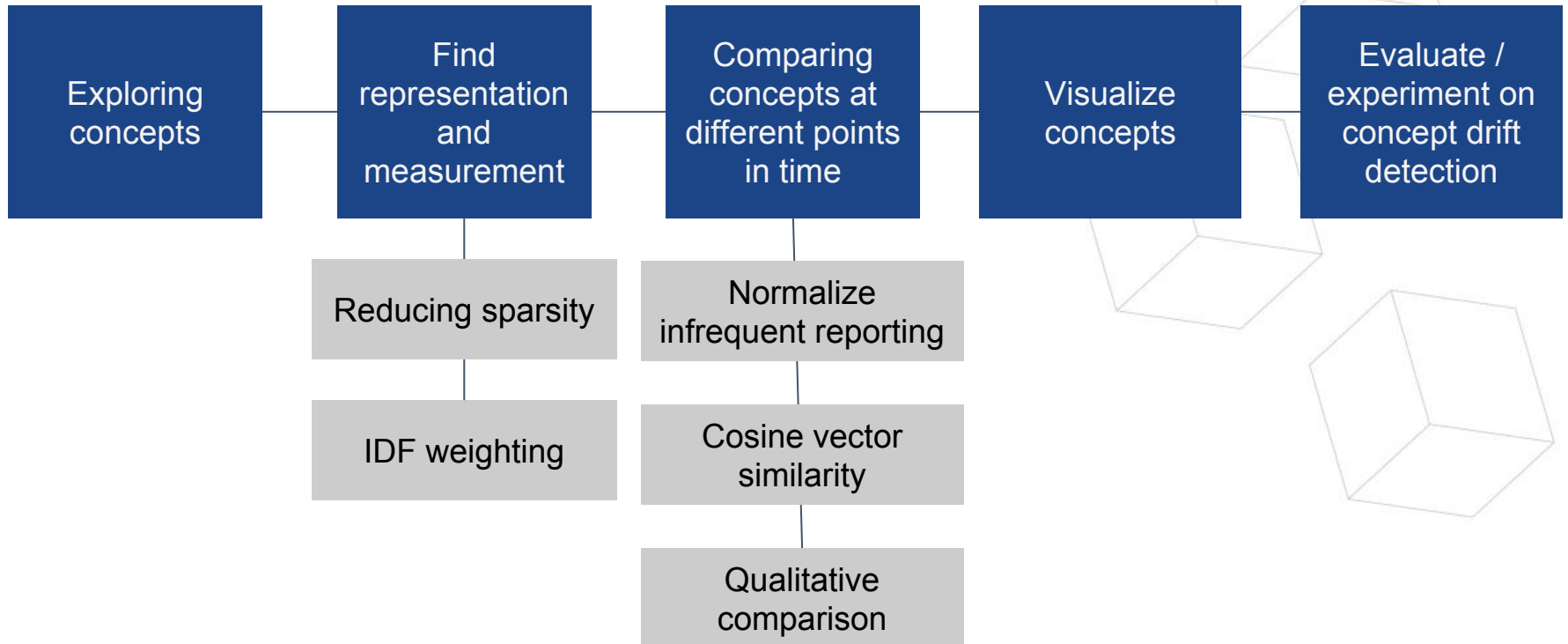
wiki/Hillary_Clinton

7	3	5	8	1	0
---	---	---	---	---	---

wiki/ = <http://en.wikipedia.org/wiki/>

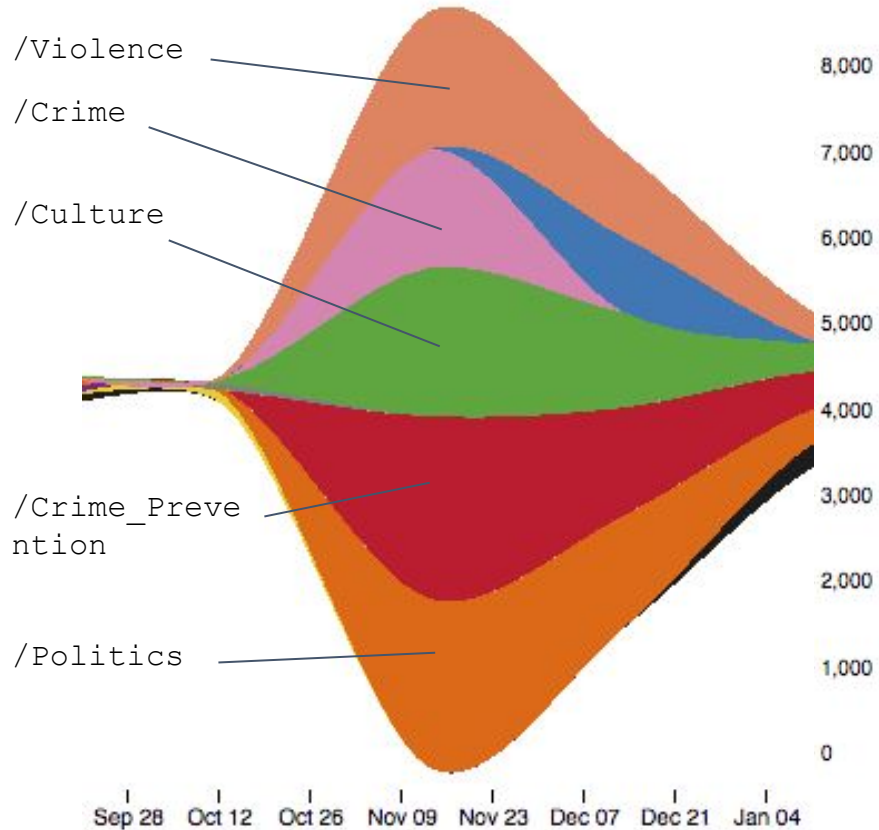
General Approach

Qualitative exploration phase



Concept: Death_Of_Eric_Garner

Core of the concept:



New context of the concept:

