

Chronicle 2050: Automatic extraction and aggregation of expected events from news articles

Level	Master
Prerequisites	Good level of Python (or optionally R)
Category	Natural Language Processing, Economics
Supervisor	Adam Jatowt

In business, politics and society, it is essential to be able to recognize trends and expected events at an early stage (including the probability of their occurrence). The aim of this topic is to make forecasts about the future easier to understand, interpret and discuss. For this purpose, knowledge about future events and general forecasts from websites should be identified, categorized thematically and chronologically, linked, and stored with the help of current methods of machine learning and natural language processing. In this way, a database of future events should be built up, which should serve as the basis for a chronicle or a calendar that covers the coming decades in relation to a user query. The chronicle is to be made available in the form of a website and serve as a basis for online discussions.

The focus of the work is the automatic extraction of events from text documents. While previous approaches were primarily aimed at current events and events from past decades, the extraction of events that are indicated in the text for the future brings additional challenges, such as the classification of plausibility (probability of occurrence) and the grouping of the same or similar events.

What should you bring along?

Basic programming knowledge.

Interest in forecasting topics, trend analysis and future research.

How is the work implemented and supervised? The work is supervised by Prof. Adam Jatowt, and will be done in collaboration with Prof. [Michael Faerber](#) at the KIT Institute AIFB, and the Innovation Hub of Atruvia AG, the IT service provider of the cooperative financial group. Financial compensation via Atruvia AG is possible.

A related work:

Adam Jatowt, Ching-man Au Yeung: Extracting collective expectations about the future from large text collections. CIKM 2011: 1259-1264