Automated Concept Map Extraction from Text

Martina Galletti[1,2,*], Inès Blin[1,3,*], Eleni Ilkou[4]

[1]Sony Computer Science Laboratories - Paris, Paris, France
[2]Sapienza University of Rome, Rome, Italy
[3]Vrije Universiteit Amsterdam, Amsterdam, The Netherlands
[4]L3S Research Center, Leibniz University, Hannover, Germany

martina.galletti @sony.com, ines.blin@sony.com, ilkou@l3s.de

LANGUAGE, DATA and KNOWLEDGE 2025

Sony CSL

VU

L3S AI Research

Context

Concept Map: semantic graph summary of concepts and their relations

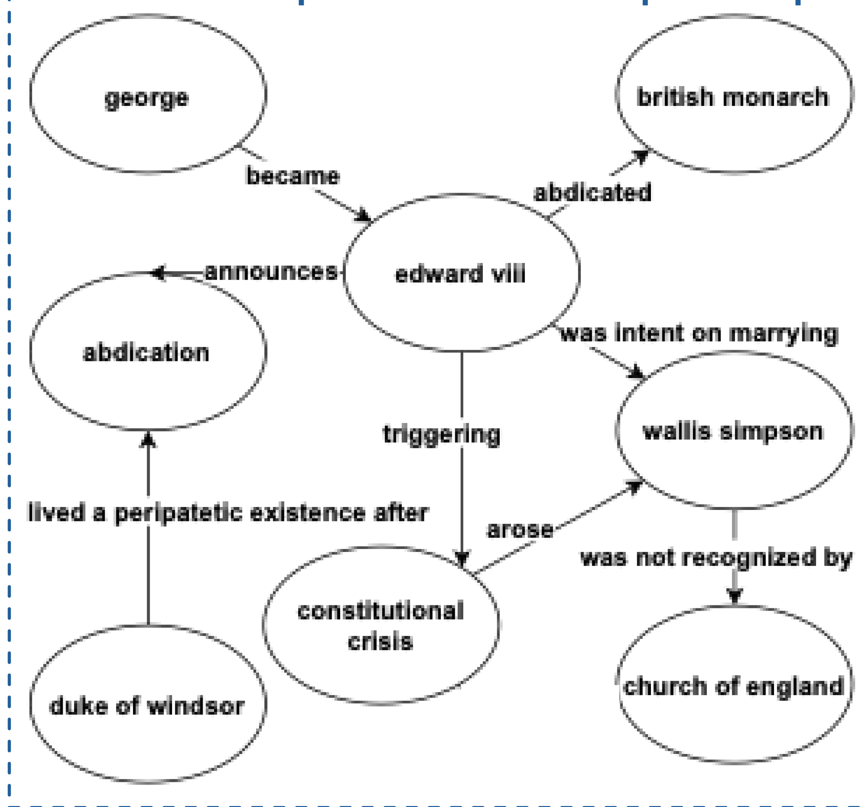
Useful for: learning new information, active learning, memory retention

- Creating them manually is **time-consuming**
- Creating them automatically is **limited** and **outdated**

We propose two novel types of approaches:

- Neuro-symbolic pipeline
- LLM-based methods

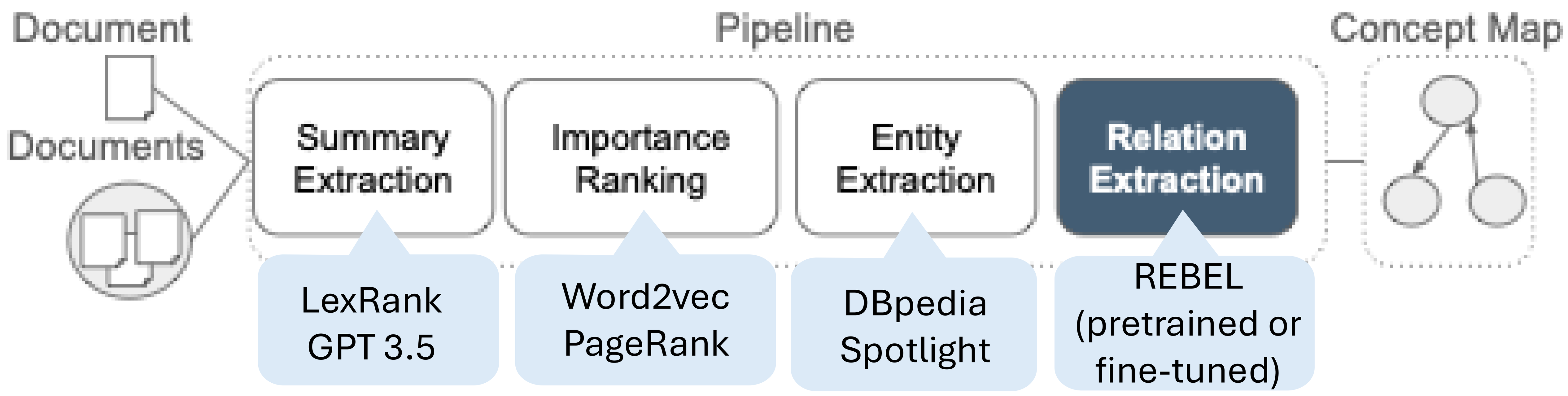
An example of Concept Map



Two novel approaches for automated concept map extraction from text

Neuro-symbolic pipeline

Overview of the pipeline



- Innovative integration of these tools within a cohesive framework for concept map extraction
- First ones to propose the summarization step as a first step

LLM-based methods

Zero-shot and One-shot

- Zero-shot: task and output description
- One-shot: Zero-shot + one example

Decomposed Prompting

Input document(s) → Summarisation → Extract entities → Group entities → Extract relations → Group relations → Most important triples → Concept Map

Experiments and Results

Experimental Setup

Datasets: WIKI [1] (main, hyperparameter search), BIOLOGY [2] (fine-tuning REBEL)

Metrics: METEOR, ROUGE-2

Baselines: from the literature

Our methods: pipeline (varying elements), LLM

Results

State-of-the-art ROUGE-2: Precision (+165%), F1 (+59%)

Ours: METEOR > ROUGE-2 (semantic quality vs. exact overlap)

- All is better
- Struggles more with relation extraction

- Decomposed is better
- State-of-the-art METEOR F1 (+48%), ROUGE-2 Recall (+4.7%)

Method	METEOR			ROUGE-2		
	Pr	Re	F1	Pr	Re	F1
Falke et al (2017)	19.6	19.0	19.2	**17.0**	10.7	**12.9**
All	24.6	**24.5**	24.0	6.4	11.8	7.6
No ranking	35.9	20.6	25.6	2.2	22.9	3.8
No summary	36.4	16.8	22.2	1.3	24.3	2.5
Zero-shot	25.2	19.1	21.2	6.3	15.9	8.2
One-shot	25.2	19.2	21.3	6.3	15.9	8.2
Decomposed	**38.4**	23.3	**28.5**	3.9	**24.3**	6.0

Results on WIKI TEST (precision, recall, F1). Bold is the highest across methods.

[1] Falke, 2019, "Automatic Structured Text Summarization with Concept Maps"
[2] Olney et al., 2011, "Generating concept map exercises from textbooks"