

Financial Event Detection Using Semantics

Introduction

Nowadays, emerging news on economic events such as acquisitions has a substantial impact on the financial markets. Therefore, it is important to be able to automatically and accurately **identify economic events in news** in a timely manner. For this, one has to be able to process a large amount of heterogeneous sources of unstructured data in order to extract knowledge useful for guiding decision making processes. We propose a **Semantics-based Pipeline for Economic Event Detection (SPEED)**, with which we aim to extract financial events from emerging news gathered from RSS feeds and to annotate these with machine-understandable meta-data, while retaining a speed that is high enough to make real-time use possible.

Framework

Our framework is modeled as a **pipeline** which takes news messages as input and is driven by a financial **ontology** developed by domain experts, containing information extracted from Yahoo! Finance on NASDAQ-100 companies. The main **components** of the pipeline are:

- **English Tokenizer:** splits text into tokens (e.g., words)
- **Ontology Gazetteer:** links identified tokens to ontology concepts
- **Sentence Splitter:** groups the tokens into sentences
- **Part-Of-Speech Tagger:** determines parts-of-speech (e.g., nouns) for each token
- **Morphological Analyzer:** reduces tokens to lemma & affix
- **Word Group Look-Up:** combines tokens into maximal word groups
- **Word Sense Disambiguator:** determines the word sense of each word group by exploring the mutual relations between word group senses using an **adapted SSI** algorithm
- **Event Phrase Gazetteer:** scans the text for events using a list of phrases or concepts
- **Event Pattern Recognition:** adds information to events with lexico-semantic patterns
- **Ontology Instantiator:** updates ontology with events

Implementation

The framework has been implemented as a **Java** application using the **General Architecture for Text Engineering (GATE)** software. We used some default GATE components, e.g., the English Tokenizer, the Sentence Splitter, the Part-Of-Speech Tagger, and the Morphological Analyzer, which generally suit our needs. Furthermore, we extended the functionality of some other GATE components (e.g., ontology gazetteering), and also implemented additional components to tackle the disambiguation and event detection processes.

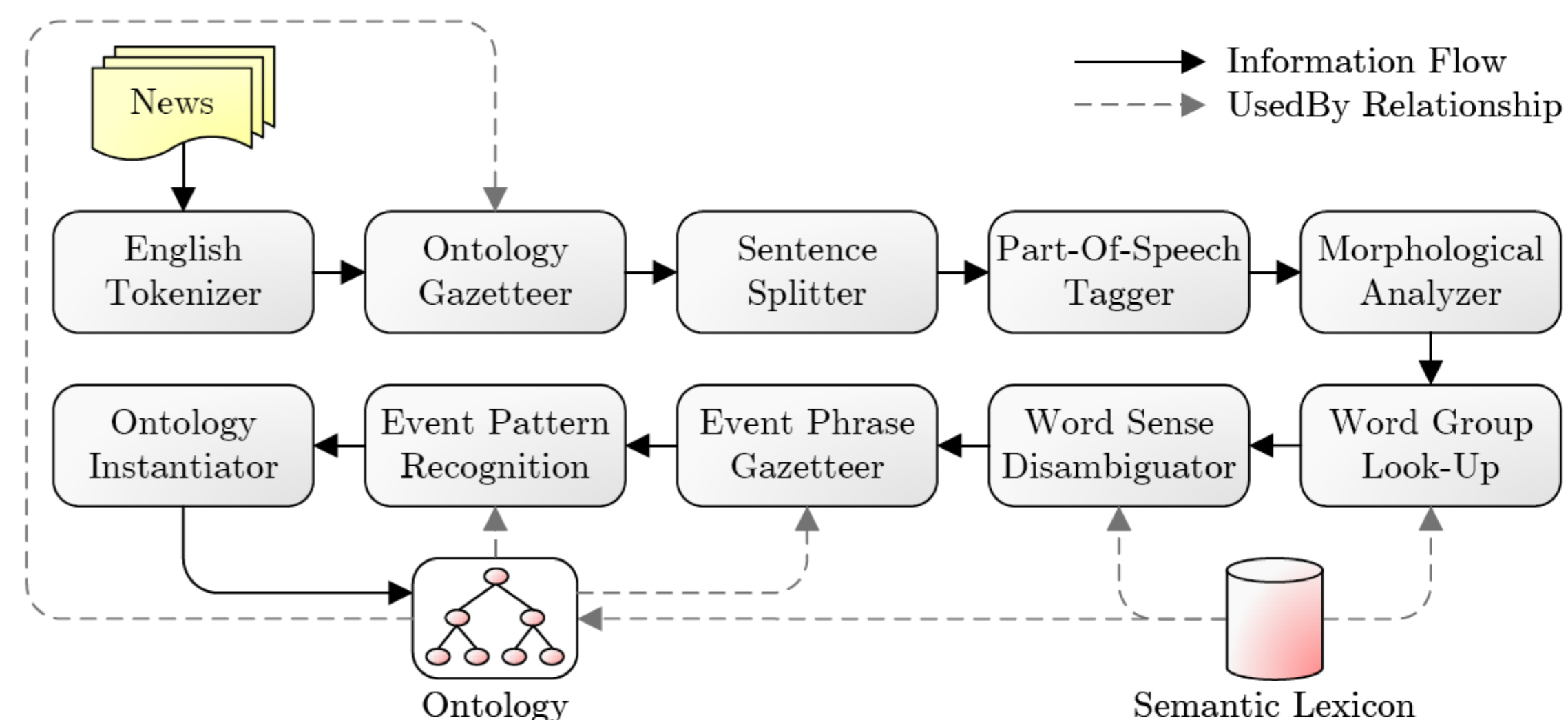


Fig. 1. SPEED design.

Results

Experiments on 200 news messages fetched from the Yahoo! Business and Technology RSS feeds show **fast** gazetteering within 632 milliseconds and a **precision** and **recall** for concept identification in news items of **86%** and **81%**, respectively. Precision and recall of **fully decorated events** result in lower values of approximately **62%** and **53%**, as they rely on multiple concepts that have to be identified correctly. Our Word Sense Disambiguator with the **adapted SSI** algorithm shows a precision and recall of **59%**, compared to a precision and recall of **53%** and **31%**, respectively, for the **original SSI** algorithm.

Conclusions

We have proposed a semantics-based framework for economic event detection (SPEED), which extracts financial events from news articles. For our implementation, we have reused existing components and developed new ones such as gazetteers and a word sense disambiguator. Also, we make use of semantic lexicons and ontologies. The framework shows high precision and recall scores in our evaluation on news feeds.

Acknowledgement

This publication was partially supported by the Dutch Organization for Scientific Research (NWO) Physical Sciences Free Competition Project 612.001.009: Financial Events Recognition in News for Algorithmic Trading (FERNAT) and the Dutch national program COMMIT.

Contact

Frederik Hogenboom
 Postal: Erasmus University Rotterdam
 P.O. Box 1738
 NL-3000 DR Rotterdam
 The Netherlands
 Phone: +31 (0)10 408 8907
 E-Mail: fhogenboom@ese.eur.nl
 Web: <http://people.few.eur.nl/fhogenboom/>

