

UTEMPL - The UIMA based Text Mining Pipeline

Jannik Strötgen, Verena Meyer, Roman Klinger and Juliane Fluck
Fraunhofer Institute for Algorithms and Scientific Computing
Schloss Birlinghoven, 53754, Sankt Augustin, Germany



Fraunhofer Institute
Algorithms and
Scientific Computing

Overview

UTEMPL is a flexible tool for information extraction and text mining:

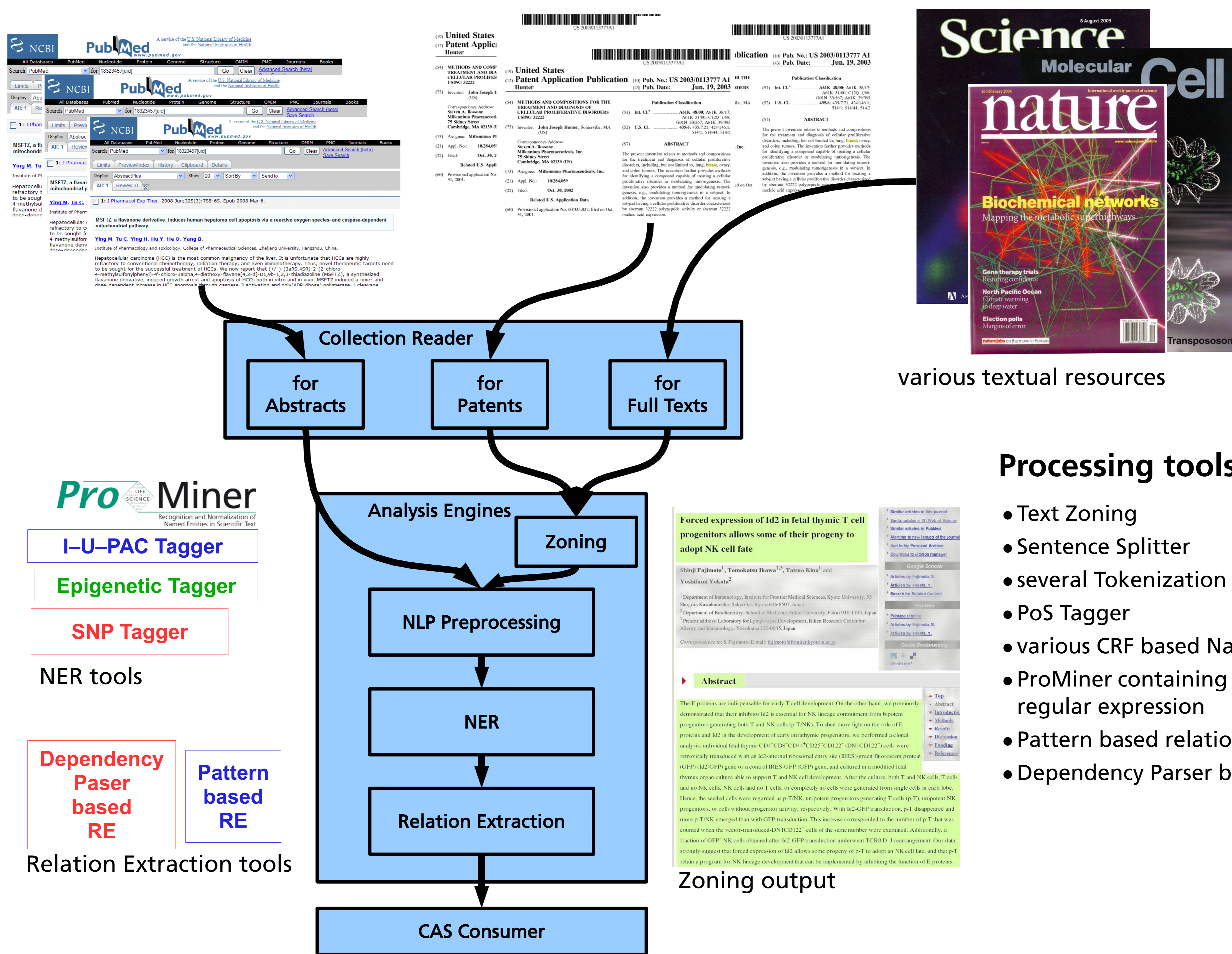
- based on UIMA
- easy use of several input formats:
 - Medline abstracts
 - Journal articles (HTML or XML)
 - Patents
 - ...

- standardized interfaces: integration of
 - NLP tools from different developers
 - new analysis engines for new tasks
- high exchangeability
- combination and selection of various internal and external applications

Advantages

- easily adaptable to new tasks
- allows to compare and evaluate different tools or methods
- efficient adaptation of existing workflows

UTEMPL workflow



Toggle Abstracts | Select All Entity Classes | Deselect All Entity Classes

Chromosomal Locations Drug Names Protein/Gene STS Marker OMIM Reference @neurIST non

Normalized SNP MeSH Disease Relations Interactions Genetic Association CRF SNP

IUPAC Cell

1. Dynamic histone H3 methylation during gene induction: **HYPB/Setd2** mediates all H3K36 trimethylation.

18157086 Authors: John W Edmunds, Louis C Mahadevan, Alison L Clayton, Date: 2008-01-23 Journal: The EMBO journal SciMag: 4.496

Understanding the function of histone modifications across inducible genes in mammalian cells requires quantitative, comparative analysis of their fate during gene activation and identification of enzymes responsible. We produced high-resolution comparative maps of the distribution and dynamics of H3K4me3, H3K36me3, H3K79me2 and H3K9ac across *Sox9* and *c-jun* upon gene induction in murine fibroblasts. In unstimulated cells, continuous turnover of H3K9 acetylation occurs on all K4-trimethylated histone H3 tails; distribution of both modifications coincides across promoter and 5' part of the coding region. In contrast, K36- and K79-methylated H3 tails, which are not dynamically acetylated, are restricted to the coding regions of these genes. Upon stimulation, transcription-dependent increases in H3K4 and H3K36 trimethylation are seen across coding regions, peaking at 5' and 3' ends, respectively. Addressing molecular mechanisms involved, we find that **Huntingtin-interacting protein HYPB/Setd2** is responsible for virtually all global and transcription-dependent H3K36 trimethylation, but not H3K36-mono- or dimethylation, in these cells. These studies reveal four distinct layers of histone modification across inducible mammalian genes and show that **HYPB/Setd2** is responsible for H3K36 trimethylation throughout the mouse nucleus.

Toggle Abstracts | Select All Entity Classes | Deselect All Entity Classes

Chromosomal Locations Drug Names Protein/Gene STS Marker OMIM Reference @neurIST non

Normalized SNP MeSH Disease Relations Interactions Genetic Association CRF SNP

IUPAC Cell

1. **Sulindac** suppresses **beta-catenin** expression in human cancer cells.

18291382 Authors: Anjia Han, Zibo Song, Chang Tong, Dong Hu, Xiuli Bi, Leonard H Augenlicht, Wancai Yang, Date: 2008-03-31 Journal: European journal of pharmacology SciMag: 0.322

Sulindac has been reported to be effective in suppressing tumor growth through the induction of **p21WAF1/cip1** in human, animal models of colon cancer and colon cancer cells. In this study, we treated human breast cancer cell line **MCF-7** and lung cancer cell line **A549** as well as colon cancer cell line **SW620** with **sulindac** to observe the effects of **sulindac** in other tissue sites. In all cell lines, proliferation was significantly inhibited by **sulindac** after 24 and 72 h of treatment. Apoptosis was induced by **sulindac** in both lung cancer cells and colon cancer cells but was not induced in breast cancer cells. Western blots showed that **p21** protein level were induced by **sulindac** in lung cancer cells and colon cancer cells, but not in breast cancer cells. However, the suppression of **beta-catenin**, a key mediator of Wnt signaling pathway, was seen in all three cell lines with **sulindac** administration. Further studies revealed that transcriptional activities of **beta-catenin** were significantly inhibited by **sulindac** and that the inhibition was **sulindac** dosage-dependent. The transcriptional targets of **beta-catenin**, **c-myc**, **cyclin D1** and **cdk4** were also dramatically downregulated. In conclusion, our data demonstrated that the efficacy of **sulindac** in the inhibition of cell proliferation (rather than the induction of apoptosis) might be through the suppression of **beta-catenin** pathway in human cancer cells.

References

- [1] D. Hanisch, K. Fundel, H. T. Mevissen, and R. Zimmer J. Fluck. ProMiner: Organism-specific protein name detection using approximate string matching. *Proceedings of the BioCreative Challenge Evaluation Workshop 2004*, 2004.
- [2] R. Klinger, C. Kolárik, J. Fluck, M. Hofmann-Apitius, and C. M. Friedrich. Detection of IUPAC and IUPAC-like names. *BMC Bioinformatics*, 24(13):i268-i276, 2008.
- [3] UIMA Homepage. <http://incubator.apache.org/uima/>. 2008.

Contact

Juliane Fluck
Fraunhofer SCAI
Schloss Birlinghoven
53754, Sankt Augustin,
juliane.fluck@scai.fraunhofer.de
<http://www.scai.fraunhofer.de>