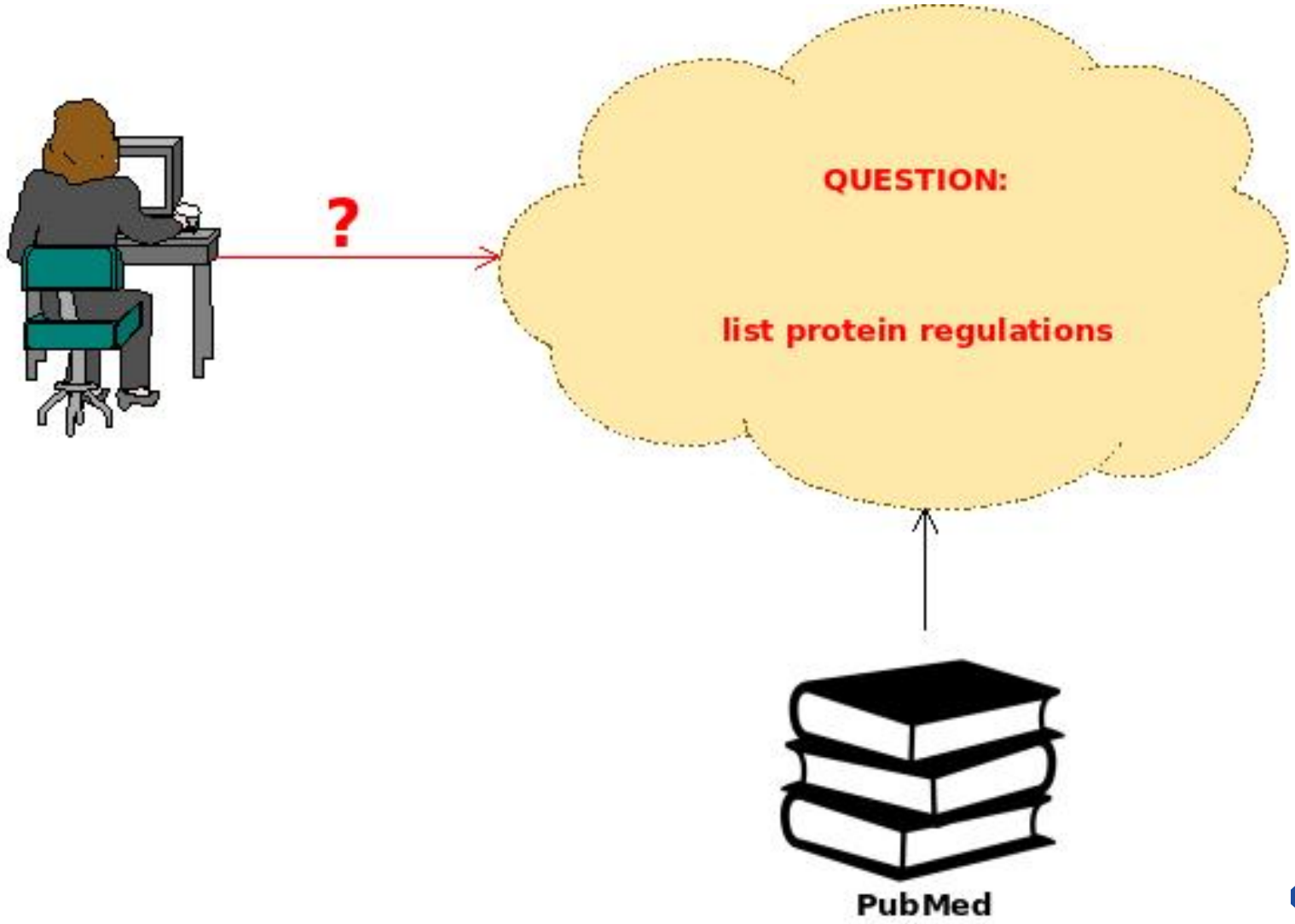
# Relevant Text Snippet Detection

IMS, Universität Stuttgart
Stuttgart, Germany
camilo.thorne@ims.uni-stuttgart.de

Abbvie, Ludwigshafen, 12.6.2017

Institut für Maschinelle Sprachverarbeitung

# Problem - Relevant Passage Retrieval




Institut für Maschinelle Sprachverarbeitung

# Relevant Passage Retrieval

(cf. PubMed PMID 9823588, abstract)

The cell cycle is a complex process that involves numerous regulatory proteins that direct the cell through a specific sequence of events culminating in mitosis and the production of two daughter cells. Central to this process are the cyclin-dependent kinases (cdks), which complex with the cyclin proteins. These proteins regulate the cell's progression through the stages of the cell cycle and are in turn regulated by numerous proteins, including p53, p21, p16, and cdc25. Downstream targets of cyclin-cdk complexes include pRb and E2F. The cell cycle can be altered to the advantage of many viral agents, most notably polyomaviruses, papillomaviruses, and adenoviruses. (...)

Institut für Maschinelle Sprachverarbeitung

# Relevant Passage Retrieval

(cf. PubMed PMID 9823588, abstract)

The cell cycle is a complex process that involves numerous regulatory proteins that direct the cell through a specific sequence of events culminating in mitosis and the production of two daughter cells. Central to this process are the cyclin-dependent kinases ( cdks ), which complex with the cyclin proteins. These proteins regulate the cell's progression through the stages of the cell cycle and are in turn regulated by numerous proteins, including p53, p21, p16, and cdc25. Downstream targets of cyclin-cdk complexes include pRb and E2F. The cell cycle can be altered to the advantage of many viral agents, most notably polyomaviruses, papillomaviruses, and adenoviruses. (...)

Entities : Proteins

Institut für Maschinelle Sprachverarbeitung

# Relevant Passage Retrieval

(cf. PubMed PMID 9823588, abstract)

The cell cycle is a complex process that involves numerous regulatory proteins that direct the cell through a specific sequence of events culminating in mitosis and the production of two daughter cells. Central to this process are the cyclin-dependent kinases ( cdks ), which complex with the cyclin proteins. These proteins regulate the cell's progression through the stages of the cell cycle and are in turn regulated by numerous proteins, including p53, p21, p16, and cdc25. Downstream targets of cyclin-cdk complexes include pRb and E2F. The cell cycle can be altered to the advantage of many viral agents, most notably polyomaviruses, papillomaviruses, and adenoviruses. (...)

Entities : Proteins

Events : Regulations

Institut für Maschinelle Sprachverarbeitung

# Relevant Passage Retrieval

(cf. PubMed PMID 9823588, abstract)

The cell cycle is a complex process that involves numerous regulatory proteins that direct the cell through a specific sequence of events culminating in mitosis and the production of two daughter cells. Central to this process are the cyclin-dependent kinases ( cdks ), which complex with the cyclin proteins. These proteins regulate the cell's progression through the stages of the cell cycle and are in turn regulated by numerous proteins, including p53 , p21 , p16 , and cdc25 . Downstream targets of cyclin-cdk complexes include pRb and E2F . The cell cycle can be altered to the advantage of many viral agents, most notably polyomaviruses, papillomaviruses, and adenoviruses. (...)

Entities : Proteins

Events : Regulations

Relations : p53 — regulated — p16 : 63.00

Institut für Maschinelle Sprachverarbeitung

# Relevant Passage Retrieval

(cf. PubMed PMID 9823588, abstract)

The cell cycle is a complex process that involves numerous regulatory proteins that direct the cell through a specific sequence of events culminating in mitosis and the production of two daughter cells. Central to this process are the cyclin-dependent kinases (cdks), which complex with the cyclin proteins. **These proteins regulate the cell's progression through the stages of the cell cycle and are in turn regulated by numerous proteins, including p53, p21, p16, and cdc25.** Downstream targets of cyclin-cdk complexes include pRb and E2F. The cell cycle can be altered to the advantage of many viral agents, most notably polyomaviruses, papillomaviruses, and adenoviruses. (...)

Select:

> *These proteins regulate the cell's progression through the stages of the cell cycle and are in turn regulated by numerous proteins, including p53, p21, p16, and cdc25.*

with confidence 63.00

Institut für Maschinelle Sprachverarbeitung

# Ranking by Confidence

**63.00** - These proteins regulate the cell's progression through the stages of the cell cycle and are in turn regulated by numerous proteins, including p53, p21, p16, and cdc25. (PMID 9823588)

**57.34** - Lack of mMCP-5 protein did not alter the granulation of the IL-3/IL-9-dependent mMCP-2+ MCs in the jejunal mucosa of Trichinella spiralis-infected mice. (PMID 28193842)

**32.73** - IGF-I and/or GH increased serum concentrations of IGF-I and IGFBP-3 (PMID 8776688)

⋮

Institut für Maschinelle Sprachverarbeitung

# System Architecture

?

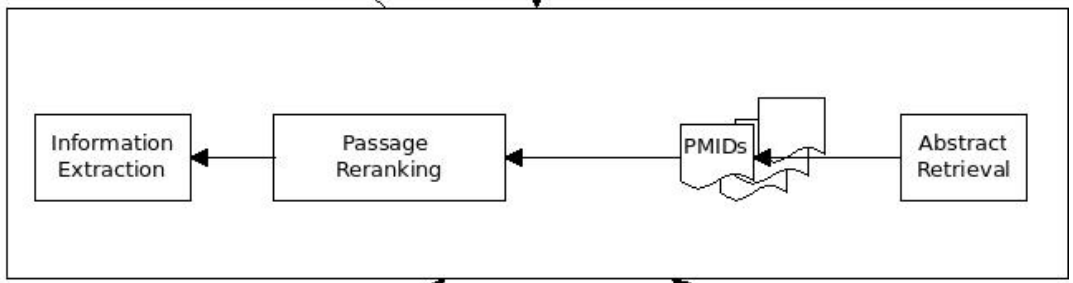
**QUESTION:**

**list protein regulations**

63.00 These proteins regulate the cell'sprogression through the stages of the cell cycle and are in turn regulated by numerous proteins, including p53, p21, p16, and cdc25.

57.34 Lack of mMCP-5 protein did not alter the granulation of the IL-3/IL-9-dependent mMCP-2+ MCs in the jejunal mucosa of Trichinella spiralis-infected mice.

...





**PubMed**

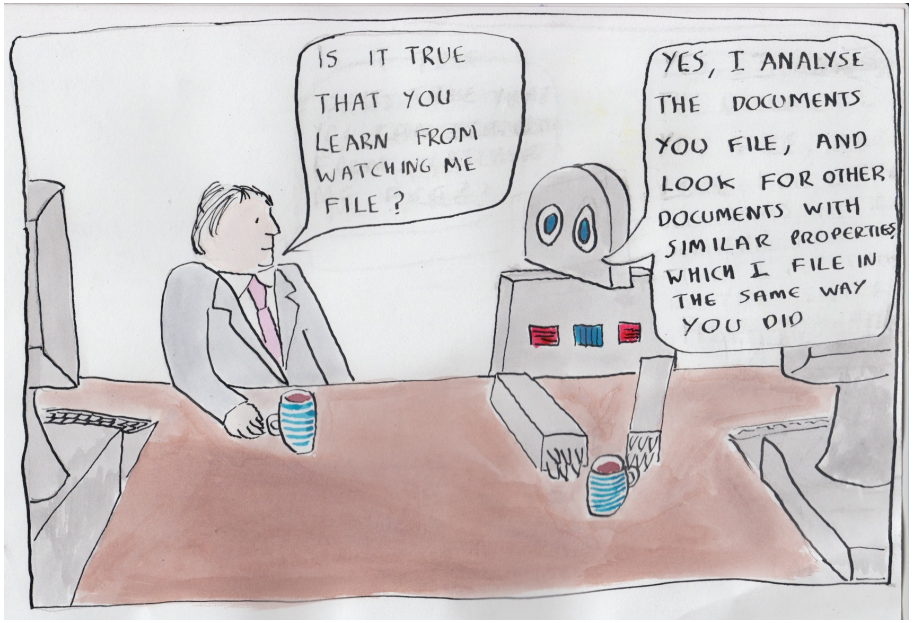
Institut für Maschinelle Sprachverarbeitung

# Confidence Estimation Models [1, 2]

1. Rely on the confidence estimates of the NER and IE systems (and weighted combinations thereof)
2. Take into account structural information from the extracted relations
   - graph-theoretical properties (from induced graph)
   - linguistic features
3. Take into account external knowledge sources (databases, ontologies, text) with different degrees of trust

Distant evaluation: apply models to MedLine papers and evaluate w.r.t. biomedical DBs

Institut für Maschinelle Sprachverarbeitung




**Thanks!**

Institut für
**Maschinelle**
**Sprachverarbeitung**

# References


[1] E. Agichtein. Confidence estimation methods for partially supervised relation extraction. In *Proc. of SIAM Intl. Conf. on Data Mining (SDM06)*. SIAM, 2006.

[2] A. Culotta and A. McCallum. Confidence estimation for information extraction. In *Proceedings of HLT-NAACL 2004: Short Papers*, HLT-NAACL-Short '04, pages 109–112, Stroudsburg, PA, USA, 2004. Association for Computational Linguistics.


Institut für Maschinelle Sprachverarbeitung