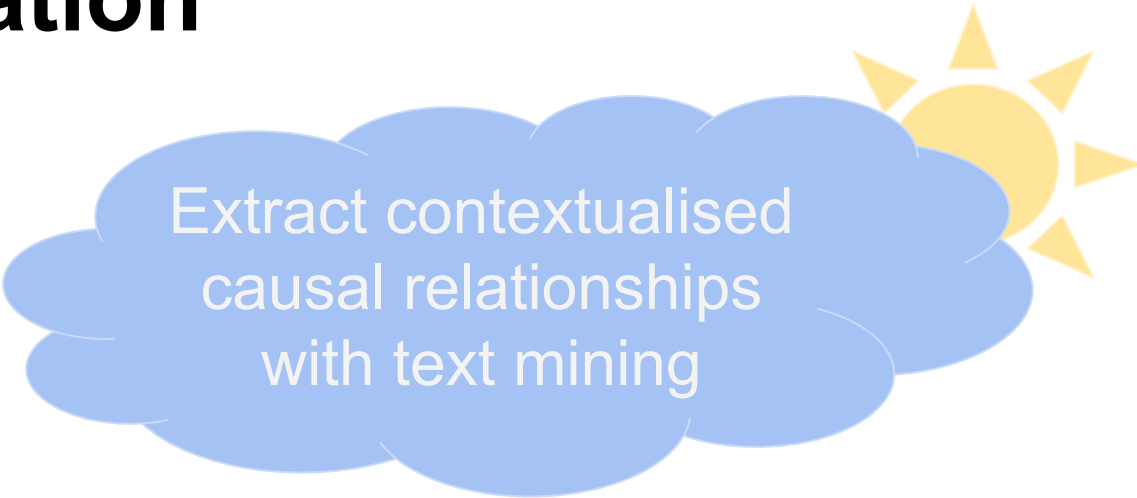


Causal statements & text mining

How can the curation of causal statements benefit from text mining?

Motivation



Extract contextualised
causal relationships
with text mining

Ease and speed up curation of causal
statements with text mining

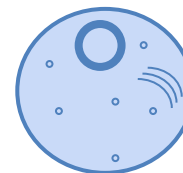
A Causal Statement



manual curation
PMID:19188143



Homo sapiens

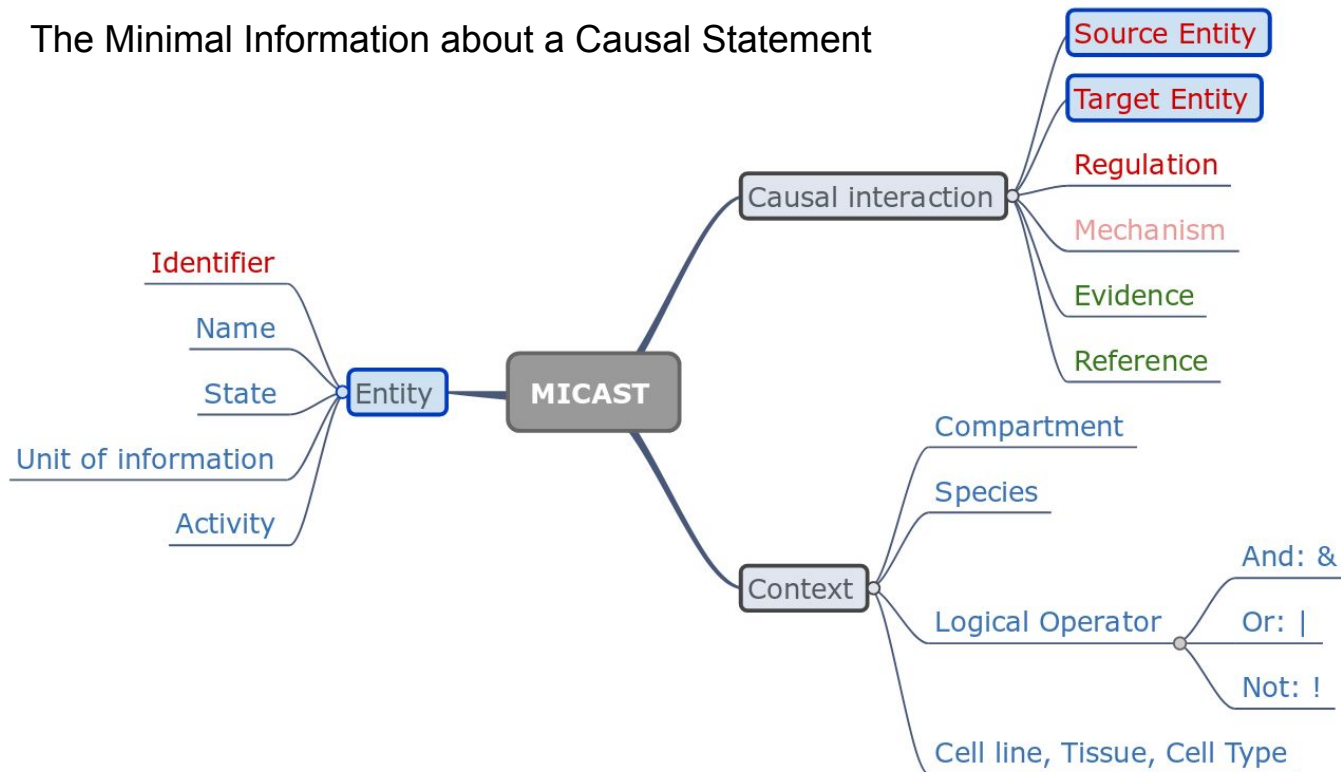


HeLa cells

Core
Metadata
Quality assessment

MICAST guideline

The Minimal Information about a Causal Statement



Core

Metadata

Quality assessment

Pre-screening steps with text mining

1. Tokenization
2. Entity recognition then mapping (genes, proteins)
3. Normalisation
4. Find verbs indicating causal relationships
5. Find biological processes or activities
6. Find evidence for:
 - a. cellular context
 - b. experiments + conditions and perturbations

Key verbs for causality

“Causes”

“is caused by”

“Influences”

“is influenced by”

“Regulates”

“is regulated by”

—————> High chance of causality in the paper

Key terms for REGULATION

Verbs

increase(s)	decrease(s)
positively	negatively regulat(ed)(es)
up-regulates	down-regulates
activat(ed)(es)	inhibit(ed)(s)
stimulat(ed)(es)	
augment(ed)s	reduc(ed)(es)
enhanc(ed)(es)	

Nouns

positive	negative regulat(ion)(or)
activation	inhibition
stimulation	
expression	repression

Key terms for MECHANISM

Mainly look for biological activities and processes

Transcription / transcriptional activity

Translation / translational activity

Modifications: phosphorylation / methylation

Processes: binding, translocation ...

And linking words

is ... by

(involved) in ... of

Example 1: ICSBP --| IFN genes

Title: Human Interferon Consensus Sequence Binding Protein Is a Negative Regulator of Enhancer Elements Common to Interferon-inducible Genes

Abstract: [...] H-ICSBP is expressed exclusively in cell lines of hematopoietic origin. The results of transient transfection assays carried out either in hematopoietic or nonhematopoietic cells suggest that ICSBP acts as a negative regulatory factor on ICs-containing promoters. Furthermore, either interferon- γ (IFN- γ) or IFN- α can alleviate the repression mediated by ICSBP. Therefore, ICSBP may be involved in maintaining submaximal transcriptional activity of IFN-inducible genes in hematopoietic cells. IFN treatment would then alleviate repression allowing maximal transcriptional activity of these genes.

[PMID:1460054](#)

Example 2: RREB-1 → CT gene

Entity

Regulation

Mechanism

Metadata

Quality

Title: RREB-1, a novel zinc finger protein, is involved in the differentiation response to Ras in human medullary thyroid carcinomas.

Abstract: This differentiation process is accompanied by a marked increase in the transcription of the human calcitonin (CT) gene.

Text: **Transcriptional activation of the CT gene by RREB-1.** The ability of RREB-1 to directly activate gene expression through its DNA binding site in the CT gene was tested with the TT:DRaf-1:ER cells.

[PMID:8816445](#)

Example 3: NZF-1 → Pit-1

Title: A Novel Family of Cys-Cys, His-Cys Zinc Finger Transcription Factors Expressed in Developing Nervous System and Pituitary Gland

Abstract: [...] This gene is expressed in a restricted fashion exhibiting distinct temporal and spatial patterns in the developing nervous system, primarily brain, spinal cord, sensory ganglia, retina, and nasal epithelia, as well as in the pituitary, and is referred to as neural zinc finger factor 1 (NZF-1). NZF-1 binds specifically to a cis-regulatory element of the β -retinoic acid receptor (RAR β) gene, as well as to other related DNA elements, including two in the upstream enhancer region of the mouse *Pit-1* gene. In heterologous cells, NZF-1 activates transcription from promoters containing specific binding sequences and can synergize with other factors, such as Pit-1, to regulate gene expression.

Discussion: [...] Although NZF-1 is potentially expressed at low levels in embryonic pituitary, significant levels of NZF-1 transcripts are present in the adult anterior pituitary gland. *In vitro*, NZF-1 specifically binds the *Pit-1* enhancer and transactivates the *Pit-1* enhancer/promoter reporter genes, making it a candidate for modulation of *Pit-1* gene regulation *in vivo*.

[PMID:8631881](#)

Challenges

Lexical and syntactic ambiguity

Causality with different linguistic expression

Buried contextual information

————→ To what extent can the MICAST concepts be identified in the text?