Recent Improvements of the BEL Information Extraction workFlow (BELIEF) for Biomedical Text Mining and Curation

Justyna Szostak[1], Sumit Madan[2], William Hayes[3], Jens Doerpinghaus[2], Juliane Fluck[2], Marja Talikka[1], Manuel C. Peitsch[1], Julia Hoeng[1]

[1] Philip Morris International R&D, Philip Morris Products S.A., Quai Jeanrenaud 5, 2000 Neuchâtel, Switzerland (Part of Philip Morris International group of companies).

[2] Fraunhofer Institute for Algorithms and Scientific Computing, Schloss Birlinghoven, 53754 Sankt Augustin, Germany

[3] Applied Dynamics Solutions LLC, Rahway NJ USA

Introduction

Construction of structured knowledge requires technology that links text mining and curation to knowledge repository. We recently presented BEL Information Extraction workFlow (BELIEF) as a tool that facilitates the transformation of unstructured information described in the literature into structured knowledge networks (1). BELIEF automatically captures causal molecular relationships from scientific text and encodes them in BEL statements. BEL (Biological Expression Language) is a computable and human readable language for representing, integrating, storing, and exchanging biological knowledge in causal and non-causal triples. Recently, we have improved the curation process by extending the biomedical vocabulary and by making the curation dashboard more flexible. Moreover, BELIEF was enhanced with the integration of the OpenBEL API that allows direct linkage to the OpenBEL platform and enables upload of curated documents into the BEL knowledge base. These technological developments of BELIEF greatly improve the curation process and make the BEL knowledge more manageable. We continually use the BELIEF to develop an extensively annotated knowledge base of BEL triples that serve as building blocks for causal biological network models.

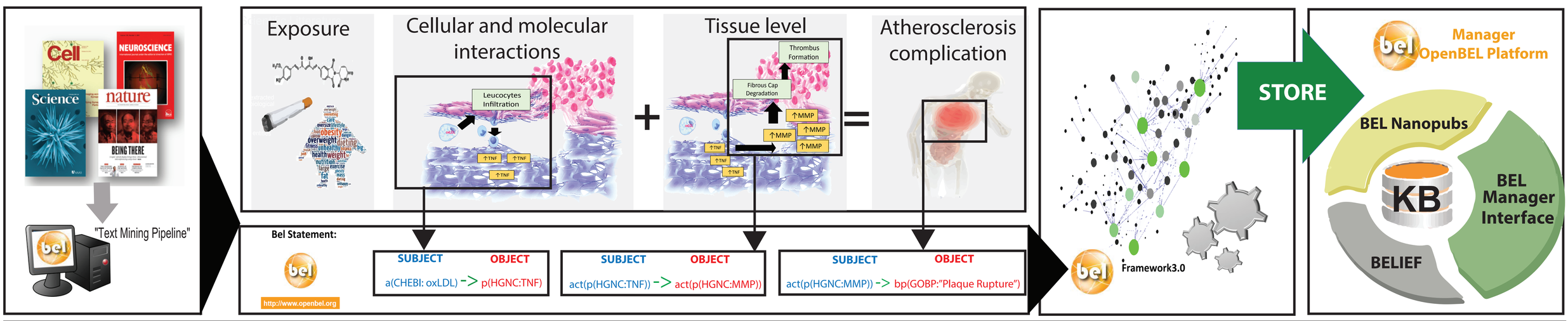
The BEL Information Extraction WorkFlow (BELIEF) Usecase

Scientific Literature

Cellular and molecular interactions taking place during disease development or progression are written into computable BEL statements

Structured Knowledge

Knowledge Repository

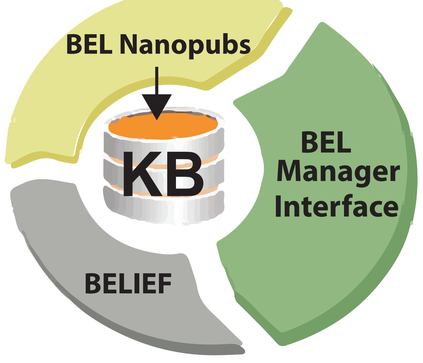


This schema describes the creation of structured knowledge base from the scientific literature. The workflow is initiated with the selection of scientific articles. After articles submission to the text mining pipeline, articles are processed, biological entities and relationships are extracted and assembled (1). The derived causal relationships describe the impact of exposure at the cellular and molecular level and also at the tissue level. The derived causal relationships are then compiled into a mechanistic network model (2). The network model describes molecular and cellular interactions that take place in vulnerable lesions accompanied with contextual information about the experiments (2).The systems toxicology approach developed by Philip Morris International (PMI) employs biological network models, transcriptomics data, and state of the art algorithms to quantify the impact of exposure on targeted biological processes (3-5).

BEL Manager : Knowledge Repository

Knowledge Repository

Manager OpenBEL Platform



BEL Nanopubs

DEFINE NAMESPACE FOR CAUSAL RELATIONSHIPS REPRESENTATIONS: 36 Define namespace category

DEFINE NAMESPACE PMICHEM AS URL "http://belief-demo.scai.fraunhofer.de/BeliefDashboard/dicten/namespaces/pmichem.belns"
DEFINE NAMESPACE GOCCID AS URL "http://resource.belframework.org/belframework/20150611/namespace/go-cellular-component-ids.belns"
DEFINE NAMESPACE MESHCS AS URL "http://resource.belframework.org/belframework/20150611/namespace/mesh-cellular-structures.belns"
DEFINE NAMESPACE DOID AS URL "http://resource.belframework.org/belframework/20150611/namespace/disease-ontology-ids.belns"
DEFINE NAMESPACE PMICOMP AS URL "http://belief-demo.scai.fraunhofer.de/BeliefDashboard/dicten/namespaces/pmicomp.belns"
DEFINE NAMESPACE DO AS URL "http://resource.belframework.org/belframework/20150611/namespace/disease-ontology.belns"
DEFINE NAMESPACE MESHPPID AS URL "http://resource.belframework.org/belframework/20150611/namespace/mesh-processes-ids.belns"

DEFINE NAMESPACE FOR CAUSAL RELATIONSHIPS ANNOTATIONS:

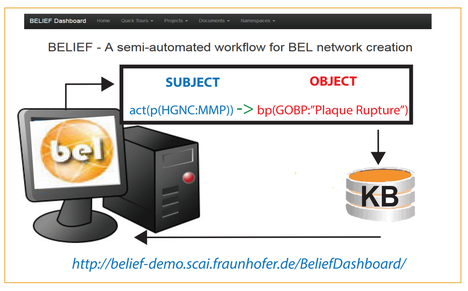
DEFINE ANNOTATION Eco AS URL "http://belief-demo.scai.fraunhofer.de/openbel/repository/annotation/evidence.belanno"
DEFINE ANNOTATION MeSHAnatomy AS URL "http://resource.belframework.org/belframework/20150611/annotation/mesh-anatomy.belanno"
DEFINE ANNOTATION SpeciesNames AS URL "http://belief-demo.scai.fraunhofer.de/openbel/repository/annotation/ncbi-taxonomy-names.belanno"
DEFINE ANNOTATION CellLine AS URL "http://resource.belframework.org/belframework/20150611/annotation/cell-line.belanno"
DEFINE ANNOTATION Disease AS URL "http://resource.belframework.org/belframework/20150611/annotation/disease.belanno"
DEFINE ANNOTATION CellStructure AS URL "http://resource.belframework.org/belframework/20150611/annotation/cell-structure.belanno"
DEFINE ANNOTATION Anatomy AS URL "http://resource.belframework.org/belframework/20150611/annotation/anatomy.belanno"
DEFINE ANNOTATION Cell AS URL "http://resource.belframework.org/belframework/20150611/annotation/cell.belanno"
DEFINE ANNOTATION Species AS URL "http://belief-demo.scai.fraunhofer.de/openbel/repository/annotation/ncbi-taxonomy-id.belanno"
DEFINE ANNOTATION MeSHDisease AS URL "http://resource.belframework.org/belframework/20150611/annotation/mesh-diseases.belanno"

BEL documents carry highly structured namespace content defined by the public database for chemicals, proteins, genes, diseases and biological processes that are employed for statements writing (1).
Additional public databases are implemented for statements annotations such as MESHAnatomy, MESHDisease, Eco...

BEL Manager Interface

The web-based BEL Manager and the underlying BEL nanopub repository provide access and allow storage of the curated relationships to the community. The BEL Manager enables the users to query and search the repository to export specific BEL contents for network refinement and reasoning.

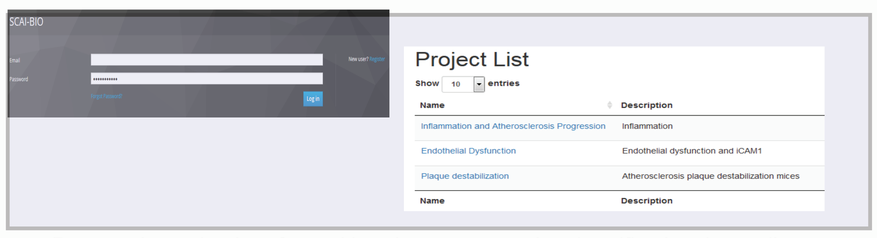
BELIEF



BELIEF (BEL Information Extraction Workflow) has recently been upgraded via the integration of the OpenBEL API enabling direct linkage to the OpenBEL
platform and allowing for direct upload of curated documents into the BEL knowledge base (KB).

BELIEF Dashboard and Functions Improvements

Integration of User Authentication and Authorization



Protected curation interface with private password authentification that redirects curators to a structured project platform.

Expansion of Biomedical Vocabulary

ZFIN.Namespace

ZFIN database were integrated in BELIEF to model mechanisms describes in Zebrafish model.

The Evidence and Conclusion Ontology (ECO) is a controlled vocabulary that describes types of scientific evidence within the area of biological research that can arise from laboratory experiments, computational methods, manual literature curation, and other means.

The NCBI Taxonomy Homepage

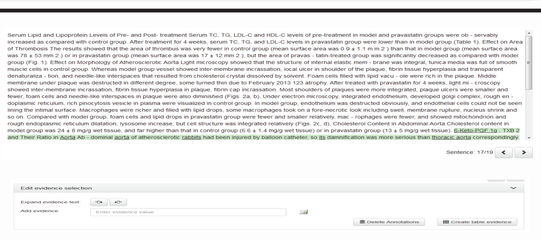
Namespace Species for annotations was extended with NCBI taxonomy. More than 19 different organisms were added to BELIEF curation.

Usability of Curation Dashboard

Full text view area

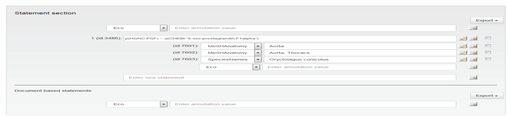
Authorised visualisation of all results part in the selected articles.

Surrounding sentences in the full-text view could be selected and add as evidence.



Statements Annotations

Flexible annotation area allowed statement annotation at the evidence level.



Search Namespace

Search concept tool box integrated into the BELIEF Dashboard, combine the curation process with searching activities on the dashboard.

BEL Nanopub Overview

BEL STATEMENT	SUBJECT: act(p(HGNC:TIMP) decrease act(p(HGNC:MMP)) — OBJECT: Relationships
CITATION	{"PubMed","PLoS One. 2012;7(10):e47134. doi: 10.1371/journal.pone.0047134. Epub 2012 Oct 11","23071737","2014-02-05","","FIXME"}
EVIDENCE	"Moreover, inhibition of MMPs by TIMP-1-overexpression resulted in decreased plaque progression, increased stabilization and decreased plaque rupture complications in murine vein grafts."
EXPERIMENTAL CONTEXT	Species: “Mus Musculus”

BEL NANOPUB

The three crucial elements of a BEL nanopub are the BEL statement showing the knowledge statement in a triple and controlled terminology, as well as the citation information and actual evidence sentence.
Experiment context is an additional field to simplify the triple assembly into biological network models (6).

Conclusions

Recent enhancements to BELIEF offer the opportunity to directly link the text mining tool and the curation process to a structured BEL Knowledge Base. A BEL nanopub represents the building block of the BEL KB. BEL Nanopub is the smallest unit of curation for BEL. We present here the integration of the concept of KB with the BELIEF text mining tool. The KB repository could be used to create, store, search and extract BEL nanopubs for network modelling or network refinement. Additional developments integrated into the BELIEF tool offer the opportunity to extend the mechanistic insight in KB and to maintain a deeply annotated contents.

References

(1) Szostak J, Ansari S, Madan S, Fluck J, Talikka M, Iskandar A et al. Construction of biological networks from unstructured information based on a semi-automated curation workflow. Database : the journal of biological databases and curation. 2015;2015:bav057. doi:10.1093/database/bav057.

(2) Szostak, J., Martin, F., Talikka, M., Peitsch, M.C., and Hoeng, J. (2016). Semi-Automated Curation Allows Causal Network Model Building for the Quantification of Age-Dependent Plaque Progression in ApoE-/- Mouse. Gene regulation and systems biology 10, 95-103.

(3) Hoeng J, Deehan R, Pratt D, Martin F, Sewer A, Thomson TM et al. A network-based approach to quantifying the impact of biologically active substances. Drug discovery today. 2012;17(9-10):413-8. doi:10.1016/j.drudis.2011.11.008.

(4) Iskandar AR, Xiang Y, Frentzel S, Talikka M, Leroy P, Kuehn D, et al. Impact assessment of cigarette smoke exposure on organotypic bronchial epithelial tissue cultures: a comparison of mono-culture and co-culture model containing fibroblasts. Toxicological Sciences. 2015:kfv122.

(5) Chindelevitch L, Ziemek D, Enayetallah A, Randhawa R, Sidders B, Brockel C et al. Causal reasoning on biological networks: interpreting transcriptional changes. Bioinformatics. 2012;28(8):1114-21. doi:10.1093/bioinformatics/bts090.

(6) Fluck, J., Madan, S., Ansari, S., Kodamullil, A.T., Karki, R., Rastegar-Mojarad, M., Catlett, N.L., Hayes, W., Szostak, J., Hoeng, J., et al. (2016). Training and evaluation corpora for the extraction of causal relationships encoded in biological expression language (BEL). Database : the journal of biological databases and curation 2016.

PMI SCIENCE PHILIP MORRIS INTERNATIONAL

This work was funded by Philip Morris International

Biocuration 2017
Bringing knowledge to the people
March 26-29, 2017 USA