

Combining resources to find answers to biomedical questions

Dina Demner-Fushman, Susanne M. Humphrey,
Nicholas C. Ide, Russell F. Loane, James G. Mork,
Miguel E. Ruiz, Patrick Ruch, Lawrence H. Smith,
W. John Wilbur, Alan R. Aronson



Task

- Find passages containing answers to biomedical questions
- 14 answer-entity types: GENES, DRUGS...
- Complex questions:
 - What [GENES] are involved in the melanogenesis of human lung cancers?
 - What [PROTEINS] in rats perform functions different from those of their human homologs?

Hypotheses

- Passage retrieval might be improved through:
 - Question “understanding”
 - Identification of potential answer entities
 - Feedback about relevance of the entire document

Hypotheses testing

- Retrieve passages with/without question processing
 - Base runs and NLMFusion
- Promote passages containing potential answers
 - LHNCBC
- Promote passages from relevant documents
 - NLMInteractive

Passage Retrieval

- Split articles into legal spans
 - Identify reference sections
- Index each span as a document
 - Essie
 - Theme
 - EasyIR
 - Indri
 - Terrier

Essie/Terrier query formulation

- Identify key entities:
 - Genes/proteins, disorders, model organisms...
- Identify answer type indicators
- Select query terms with specific semantic types and parts of speech
- Expand selected terms
- Expand answer type terms

Query example

Functional Concept v

- What [GENES] are **involved** in the melanogenesis of human lung cancers?

Molecular Function n

Human n,j

Neoplastic Process ?

- Essie: FAVOR[3.0] TOPIC[Genes] AND (human) AND (melanogenesis) AND (lung cancers)
- Terrier:
 - Topic (remove square brackets)
 - Title: melanogenesis and human and lung cancers
 - Title + Topic as description

Semantic post-processing

- SemRep re-ranking: promote passages with potential answers identified as relations
 - Glutathione S-Transferase | ASSOCIATED_WITH | Carcinoma of lung
- Semantic Filtering re-ranking: promote passages with high frequency of query terms and answer type entities
- Combine Essie, SemRep, and Semantic Filtering scores (LHNCBC run)

Relevance feedback: Manual PubMed search

What [GENES] are induced by LPS in diabetic mice?

- (lipopolysaccharides OR Ips) AND diabetes mellitus[mh] AND mice[mh] AND (gene OR genes OR ge[sh]) AND (free full text[sb])
- (("lipopolysaccharides"[MeSH Terms] OR lipopolysaccharides[Text Word]) OR Ips[All Fields]) AND "diabetes mellitus"[MeSH Terms] AND "mice"[MeSH Terms] AND (((("genes"[TIAB] NOT Medline[SB]) OR "genes"[MeSH Terms] OR gene[Text Word]) OR ("genes"[MeSH Terms] OR genes[Text Word]) OR "genetics"[Subheading]) AND "loattrfree full text"[sb])

Relevance Feedback: Document selection

Caspase-1 is not required for type 1 diabetes in the NOD mouse.

Schott WH, Haskell BD, Tse HM, Milton MJ, Piganelli JD, Choisy-Rossi CM, Reifsnyder PC, Chervonsky AV, Leiter EH.
The Jackson Laboratory, Bar Harbor, Maine 04609, USA.

Interleukin (IL)-1 beta and IL-18 are two cytokines associated with the immunopathogenesis of diabetes in NOD mice. Both of these cytokines are cleaved by caspase-1 to their biologically active forms. IL-1 is a proinflammatory cytokine linked to beta-cell damage, and IL-18 stimulates production of

interferon (IFN)gamma in synergy with IL-12. To examine the effects produced by caspase-1 deficiency on diabetes development in NOD/Lt mice, a disrupted Casp1 gene was introduced by a speed congenic technique.

Casp1(-/-) bone marrow-derived macrophages stimulated with lipopolysaccharide produced no detectable IL-18, fourfold lower IL-1 beta, and 20-30% less IL-1 alpha than macrophages from wild-type Casp1(+/+) or Casp1(+/-) controls. Unexpectedly, despite reduced IL-1 and IL-18, there was no change in the rate of diabetes or in total incidence as compared with that in wild-type NOD mice. IL-1

reportedly makes an important pathological contribution in the multidose streptozotocin model of diabetes; however, there was no difference in sensitivity to streptozotocin between NOD mice and NOD.Casp1(-/-) mice at 40 mg/kg body wt or at 25 mg/kg body wt dosage levels. These findings show that caspase-1 processing of IL-1 beta and IL-18 is not absolutely required for mediation of spontaneous or chemically induced diabetes pathogenesis in the NOD mouse.

MH- Diabetes Mellitus, Type 1/epidemiology/*genetics/physiopathology

MH- Interleukin-1/secretion

MH- Lipopolysaccharides/toxicity

MH- Mice, Inbred NOD

Relevance feedback: Passage re-ranking

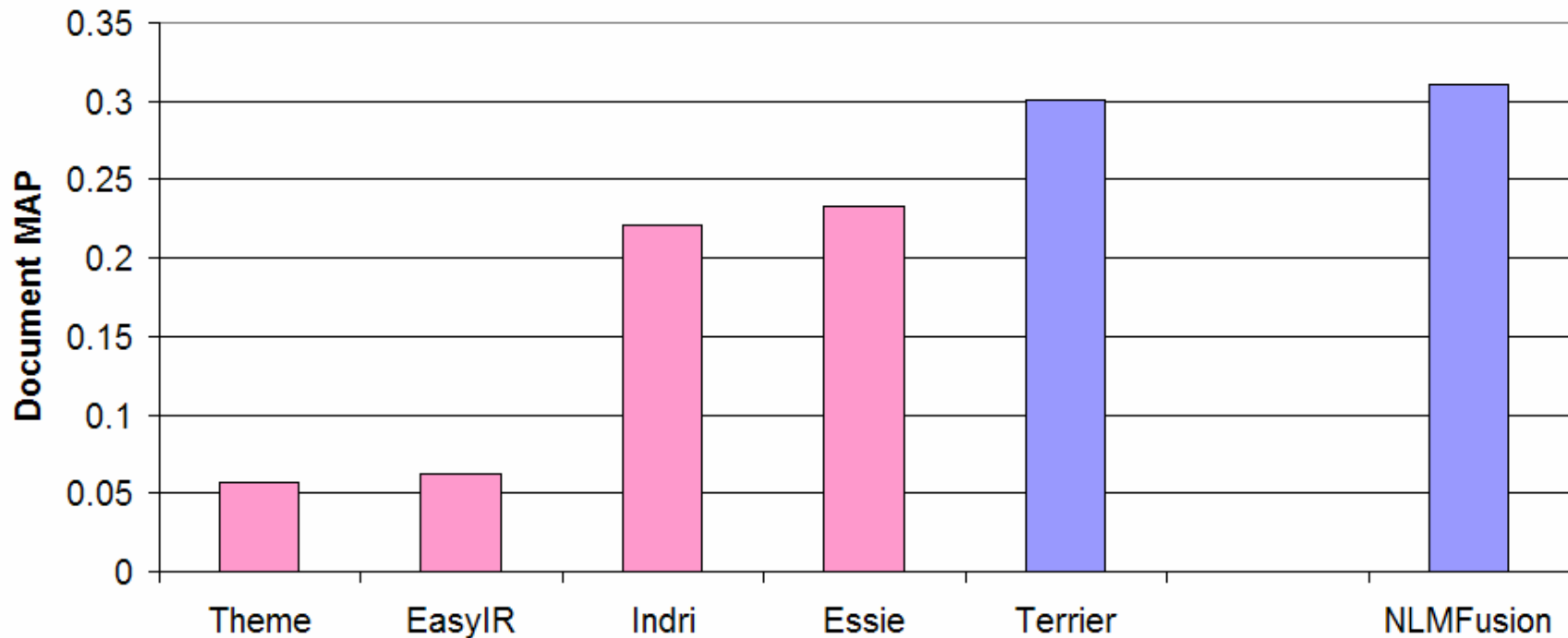
- Intersect relevant documents IDs with NLMfusion retrieval results
- Promote passages from relevant documents
 - Caspase-1 is important in the regulation of IFN production induced by lipopolysaccharide (LPS)-stimulated secretion of IL-18.... Pretreatment of NOD female mice with soluble IL-1 receptor antagonist prevented cyclophosphamide-induced hyperglycemia, but not insulinitis.

Results: Query formulation

Terrier queries	Document	Passage2	Aspect
Baseline: topics without processing	0.269	0.082	0.228
Title: ANDed identified key terms	0.297	0.089	0.250
TD: Title + Baseline as description	0.301*	0.092*	0.249

* Results significantly better than the baseline

Results: Fusion baseline



Runs significantly improved by NLMfusion

NOTE: the official EasyIR run achieved 0.199 MAP
Using this run for fusion results in 0.236 MAP

Results: Semantic processing

	Document	Passage2	Aspect
Essie (Base for re-ranking)	0.232	0.070	0.225
SemRep re-ranking	0.190	0.047	0.153
Semantic Filtering (SF) re-ranking	0.095	0.014	0.083
LHNCBC (Essie +SemRep +SF)	0.227	0.069	0.203



Re-ranking significantly outperformed by the baseline

Results: Relevance Feedback

	Document	Passage2	Aspect
NLMFusion (Base for NLMInteractive)	0.311	0.109	0.250
NLMInteractive	0.329*	0.115*	0.263

* Statistically significant difference

Results: Other

- Excluding reference sections:
 - Essie document: 0.231 vs. 0.232
 - Essie passage2: 0.051 vs. 0.070*
 - Essie aspect: 0.196 vs. 0.225
- Passage trimming:
 - LHNCBC passage1 0.0695 vs. 0.0692
 - LHNCBC passage2 0.0686 vs. 0.0679

* Statistically significant difference

Conclusions

- Question processing matters
- Fusion stabilizes results
- Relevance feedback on document level improves passage retrieval



THANK YOU FOR FIVE EXCITING YEARS!