Semantic Tagging of Scientific Articles

13th International Conference on Electronic Publishing

DC-Social Tagging community workshop
ELPUB 2009, Milan, Italy

June 10, 2009

Sudeshna Das [1,2,3]
[1]Initiative in Innovative Computing, Harvard University
[2]MIND, Massachusetts General Hospital
[3]Harvard Medical School

Content

- What is semantic tagging?
- Enabling semantic tagging
- Demo
- Use of semantic tagging

What is semantic tagging?

- Tagging with a term or resource
  - belongs to a defined class
  - that has a URI (Uniform Resource Identifier)
  - controlled vocabularies or ontologies

- The tag has meaning!
  - Defined relationship between document and tag
  - Document “discusses” PhenomenonA
  - Document “majorTopic” Gene123
  - DocumentA “cites” DocumentB

- Provenance and status
- Tag - annotation

Enabling semantic annotation

- Semi-automatic text-mining
  - Use machines to "suggest terms"
  - Use editors to refine choice
- Currently mining for
  - Gene names
  - Gene Ontology terms

Annotation of articles

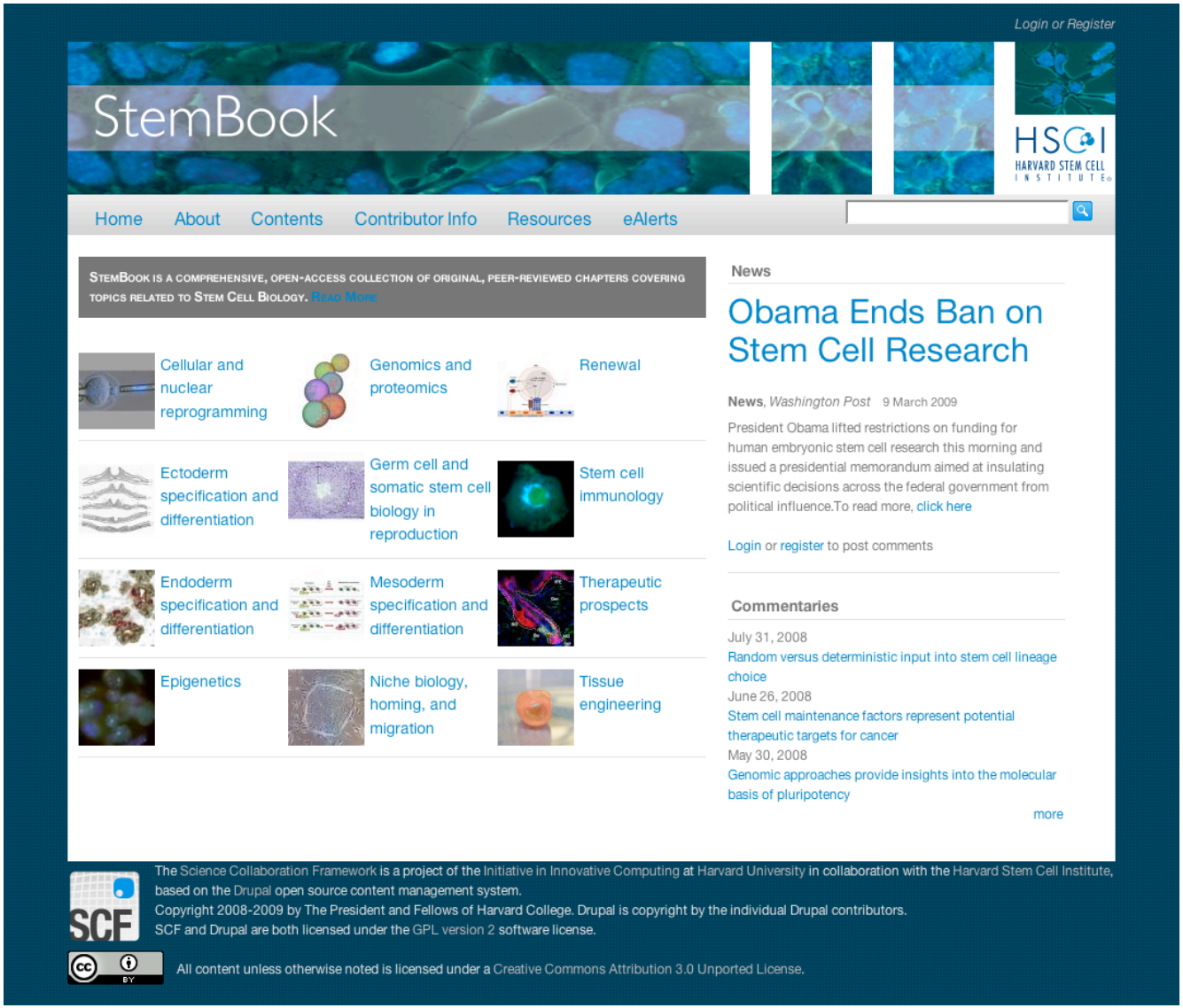
- Semi-automatic text-mining
  - Use machines to “suggest terms”
  - Use editors to refine choice
- Currently mining for
  - Gene names
  - Gene Ontology terms

✔ High recall

✔ High precision

StemBook

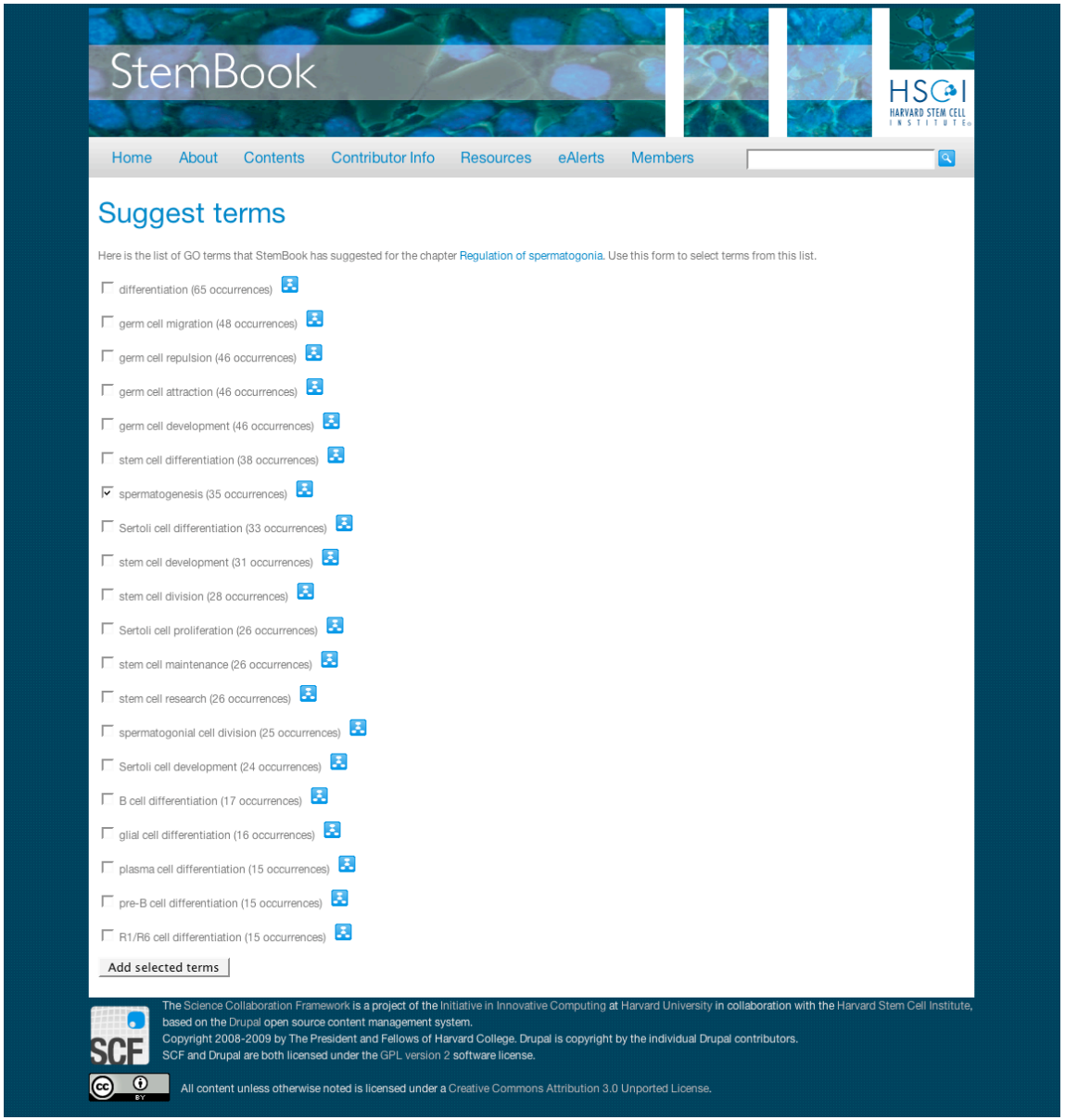
www.stembook.org



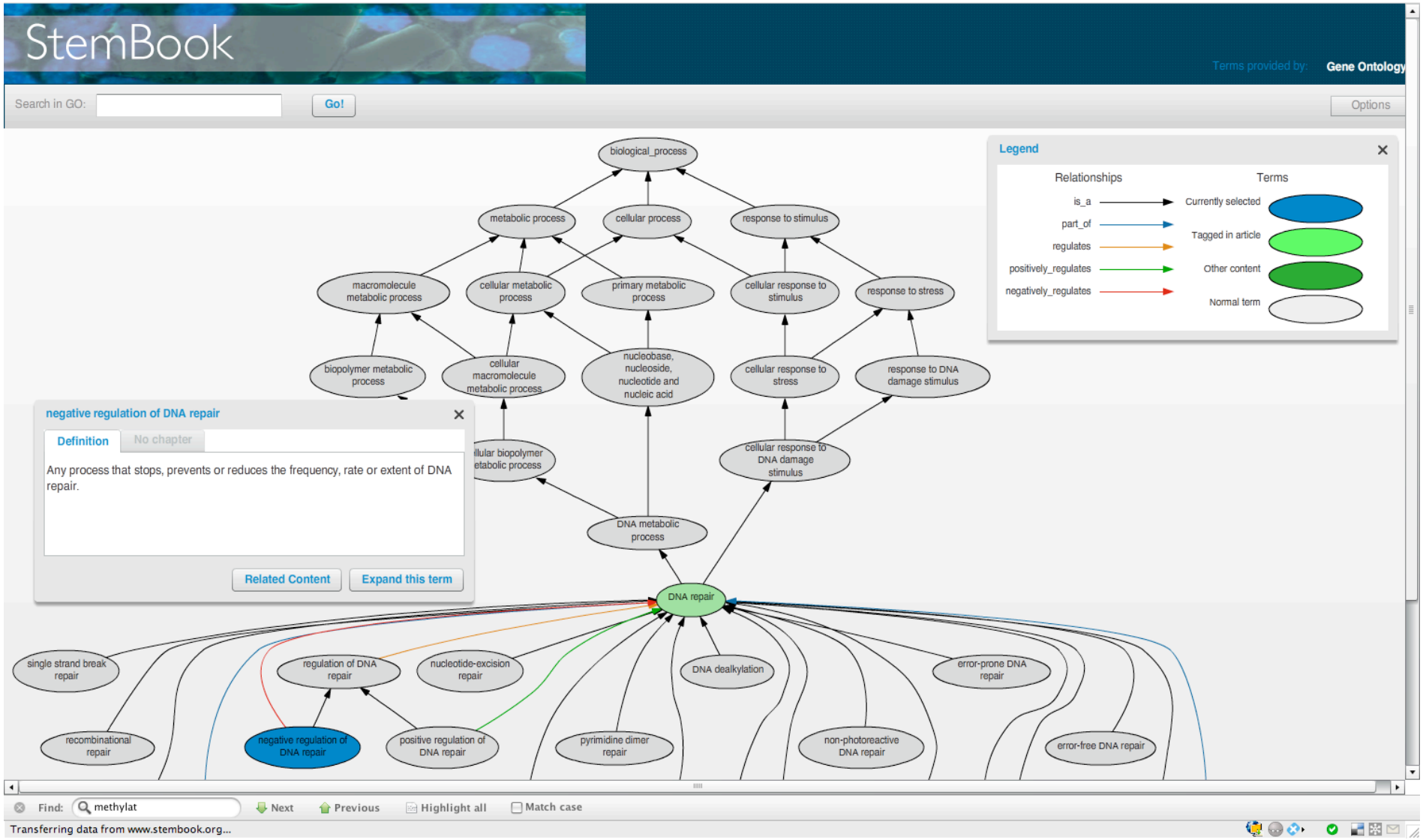
StemBook

- Online, open access, review of stem cell biology
- Collaborative effort with HSCI
  - Lisa Girard PhD (editor)
- Editorial review
  - 14 Harvard-affiliated stem cell experts and 13 non-Harvard
- Launched in Sep 2008
  - 39 articles and 89 commissioned in 12 sections
  - 14557 unique visitors
  - ~150 visitors/day

Text mining - suggested terms



Editorial review



Semi-automatic Annotation

Neuroinflammation

By Chris Scientist, Clinician, Washington University

The brain has generally been considered to be immune privileged and thus largely protected from immune factors. However, it is now recognized that the brain can initiate injury responses including neuroinflammation as a means to reduce injury and clean up injured brain tissue. The main cellular responders to brain injury are microglia, which produce a number of factors to regulate the injury response including cytokines and protective factors. Although neuroinflammation may have initial protective effects, prolonged and chronic neuroinflammation may lead to prolonged neurodegeneration such as that seen in PD. Whether neuroinflammation acts as an initial trigger of PD pathogenesis or is a downstream result of another triggering pathogenic event is unclear. Neuroinflammation may lead to increased oxidative stress. Evidence for neuroinflammation in PD includes presence of activated microglia, increased expression of cytokines and pro-inflammatory signaling cascades (e.g., NF-kB). Furthermore, epidemiological data supports a reduce risk of PD in users of anti-inflammatory drugs (e.g., NSAIDS).

References

Tansey MG, Frank-Cannon TC, McCoy MK, Lee JK, Martinez TN, McAlpine FE, Ruhn KA, Tran TA. **Neuroinflammation in Parkinson's disease: is there sufficient evidence for mechanism-based interventional therapy?** Front Biosci. 2008 Jan 1;13:709-17. PubMed

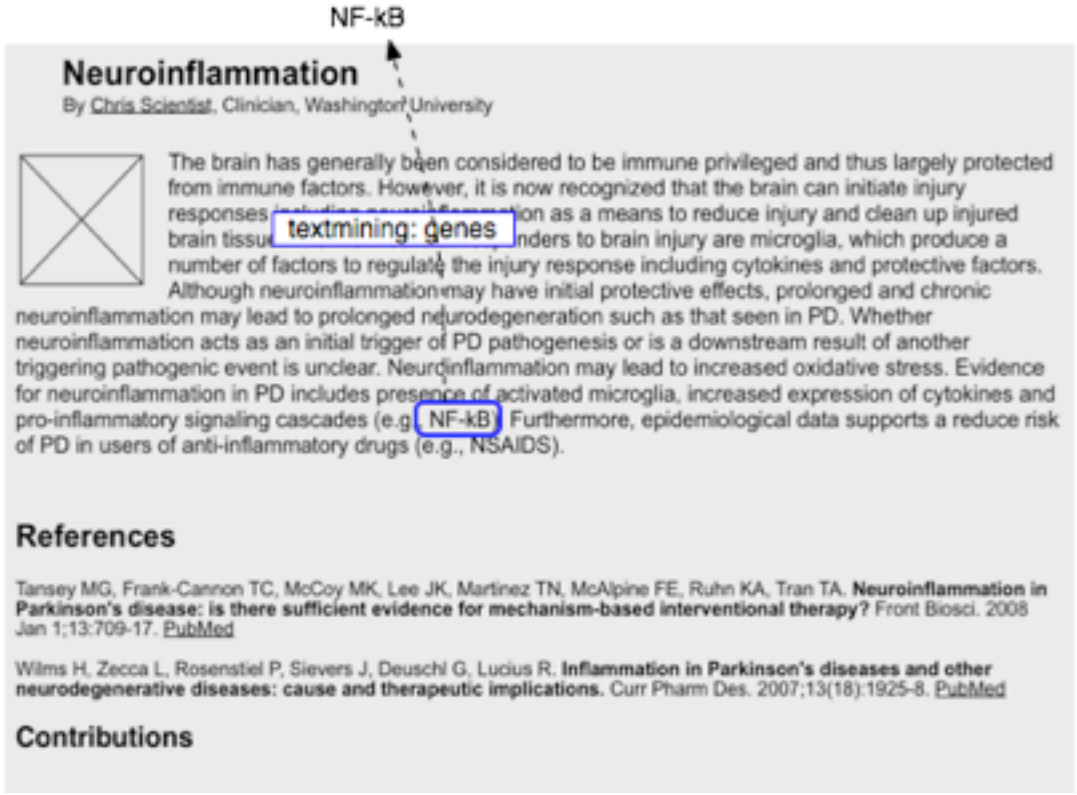
Wilms H, Zecca L, Rosenstiel P, Sievers J, Deuschl G, Lucius R. **Inflammation in Parkinson's diseases and other neurodegenerative diseases: cause and therapeutic implications.** Curr Pharm Des. 2007;13(18):1925-8. PubMed

Contributions

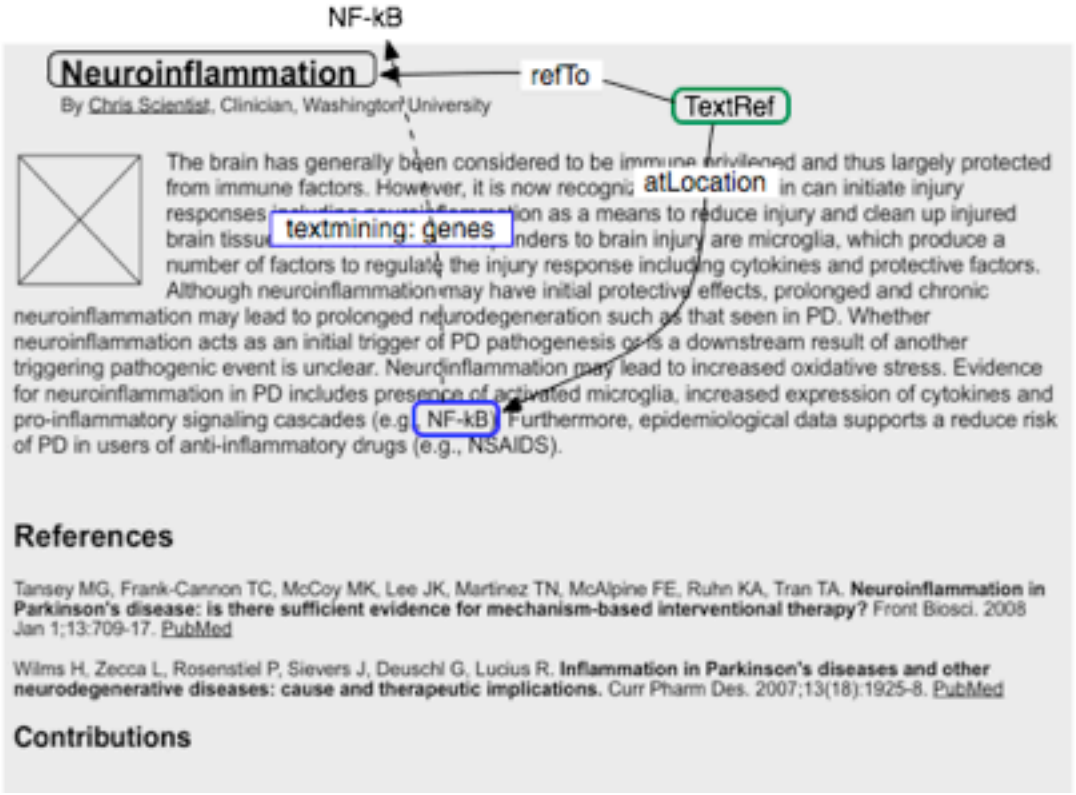
Cannabinoids and neuroprotection in basal ganglia disorders — 6 May 2008

J. Fernandez-Ruiz, Ciudad Universitaria, Madrid

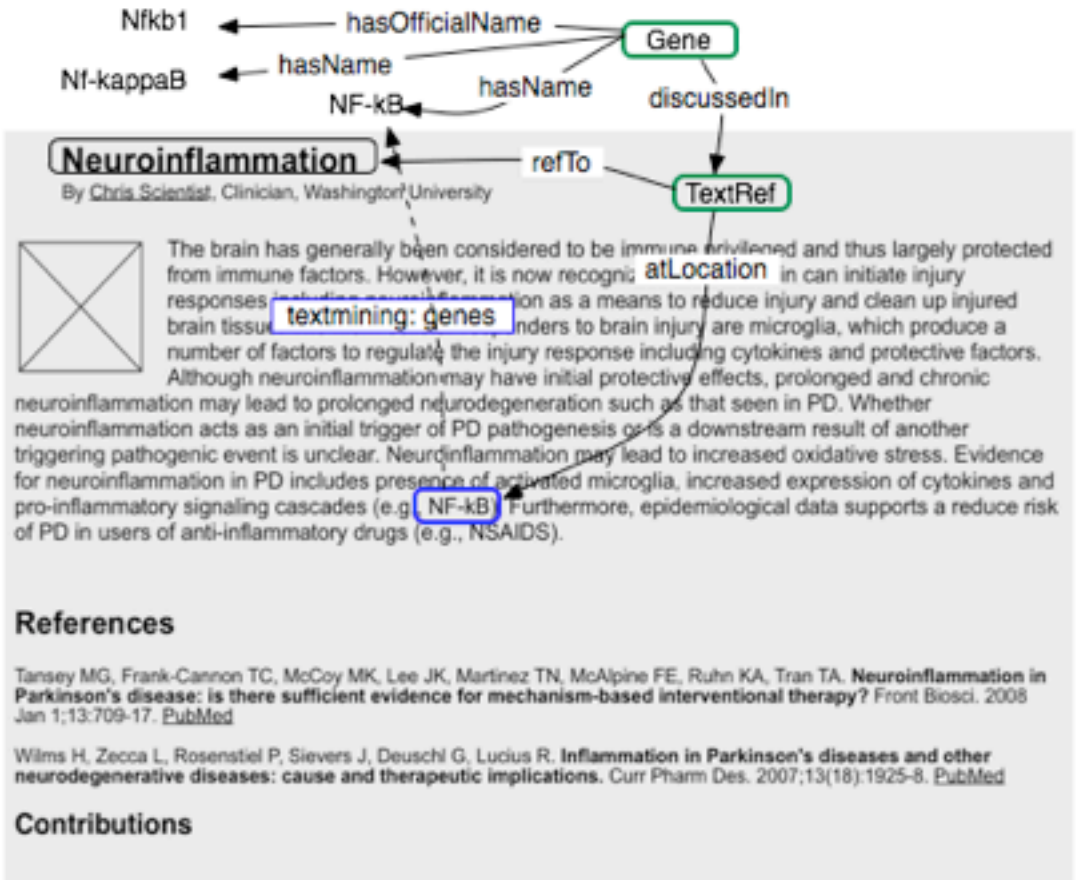
Semi-automatic Annotation



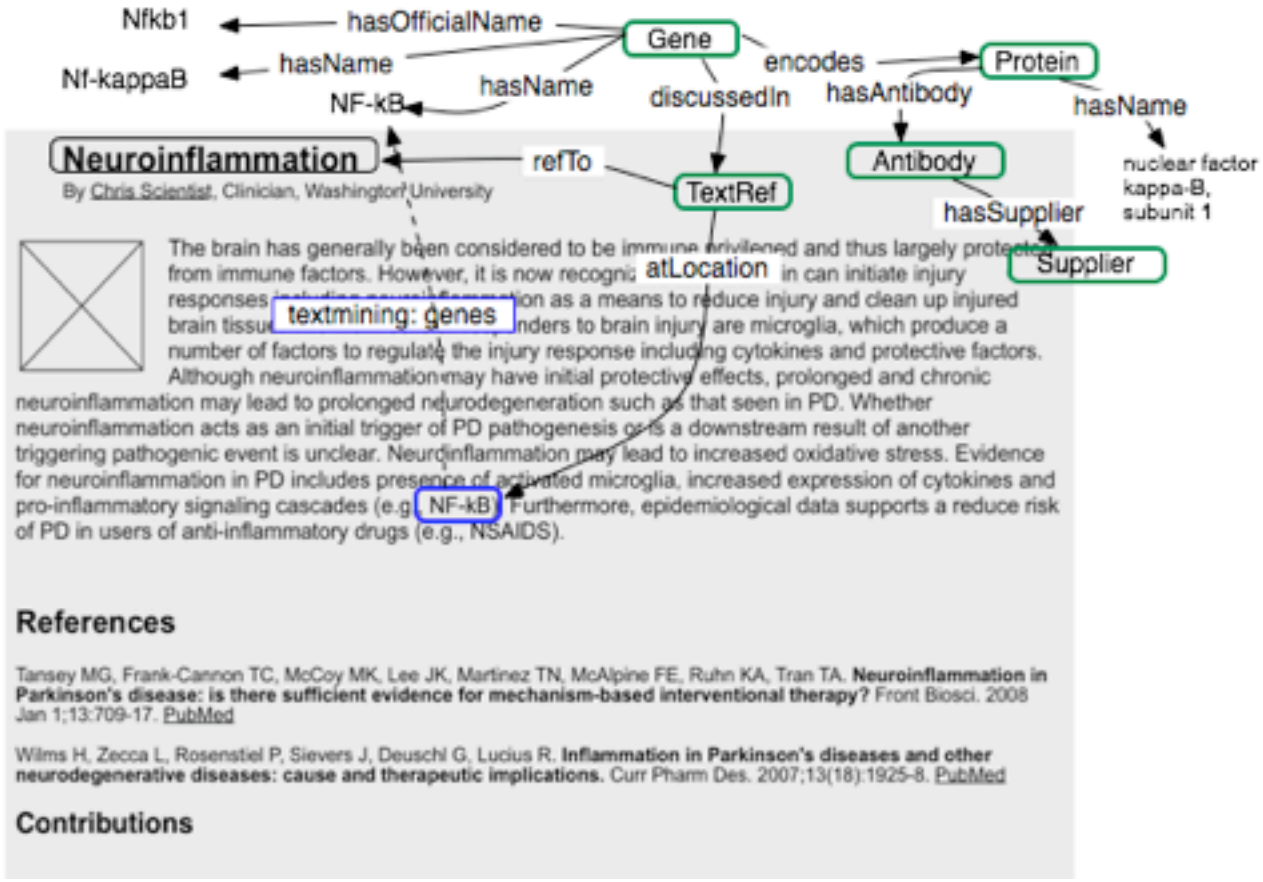
Semi-automatic Annotation



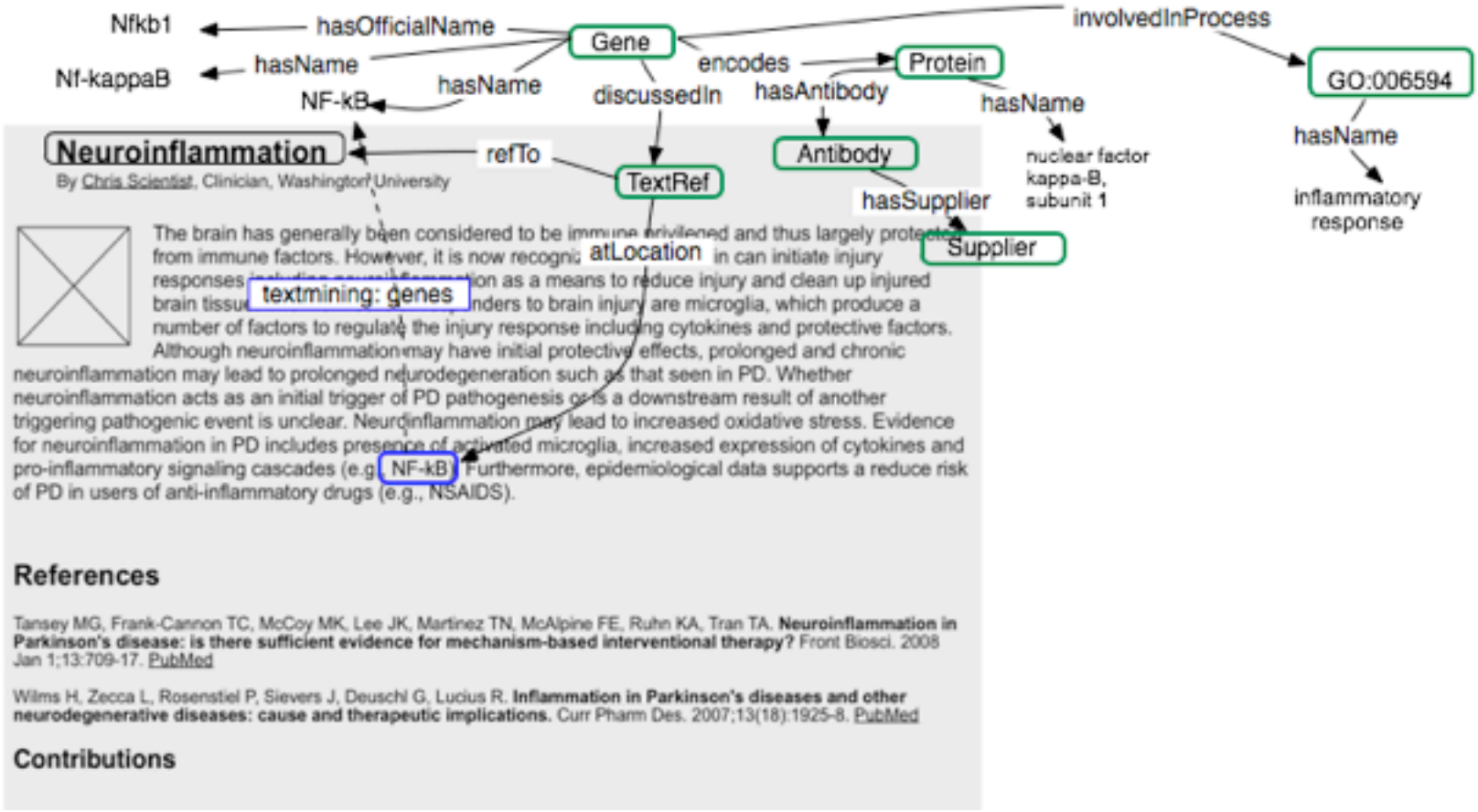
Semi-automatic Annotation



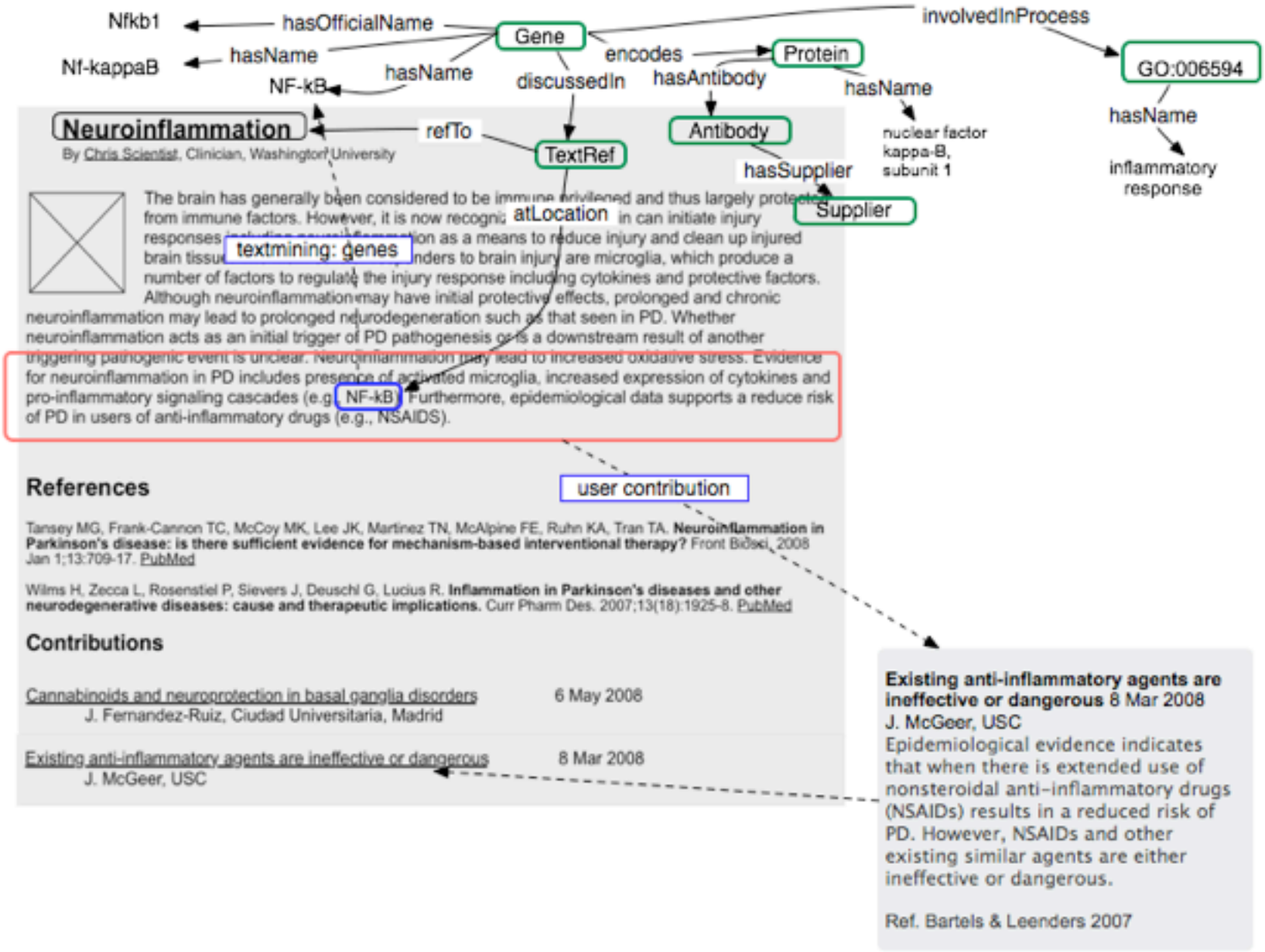
Semi-automatic Annotation



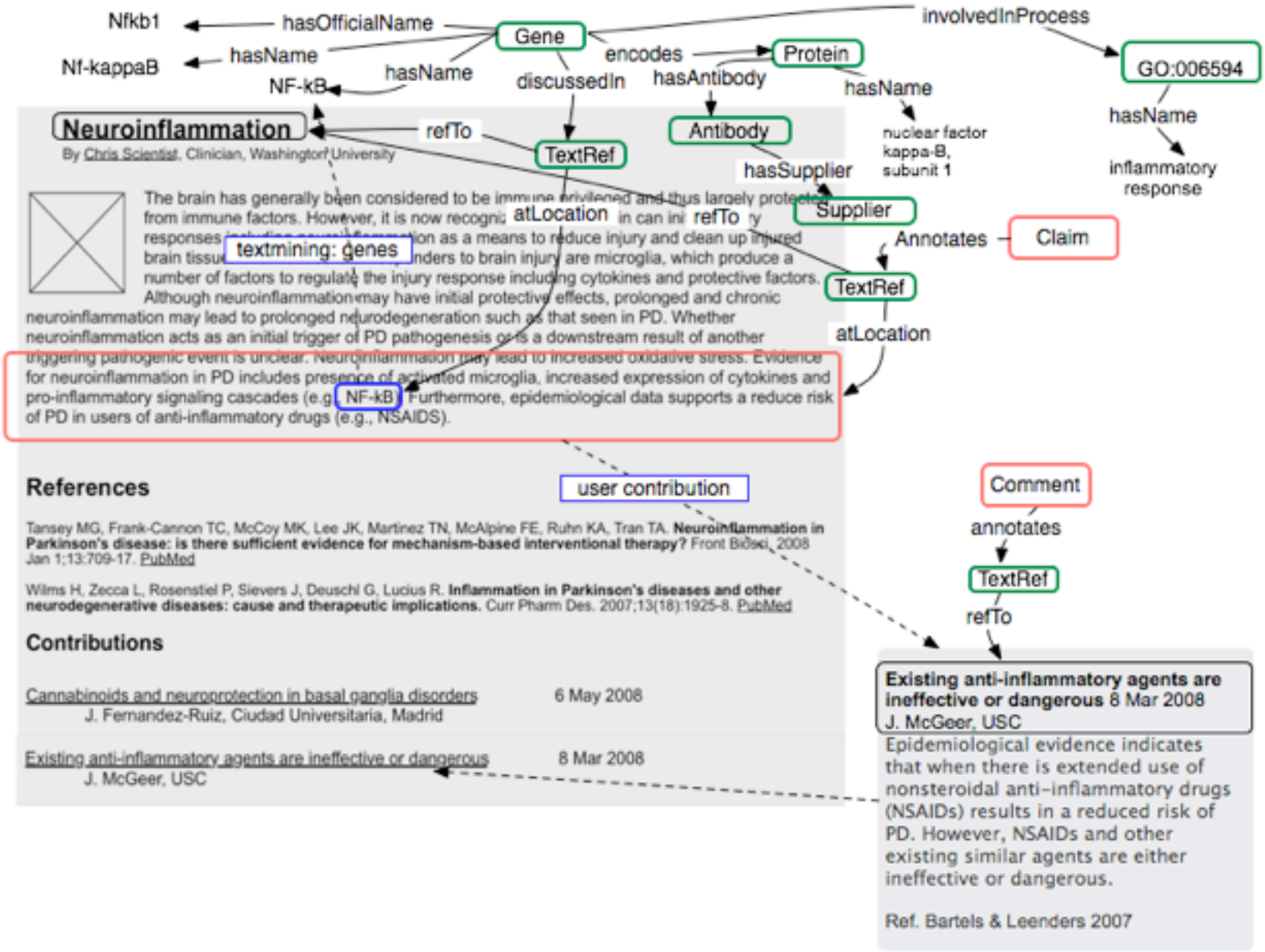
Semi-automatic Annotation



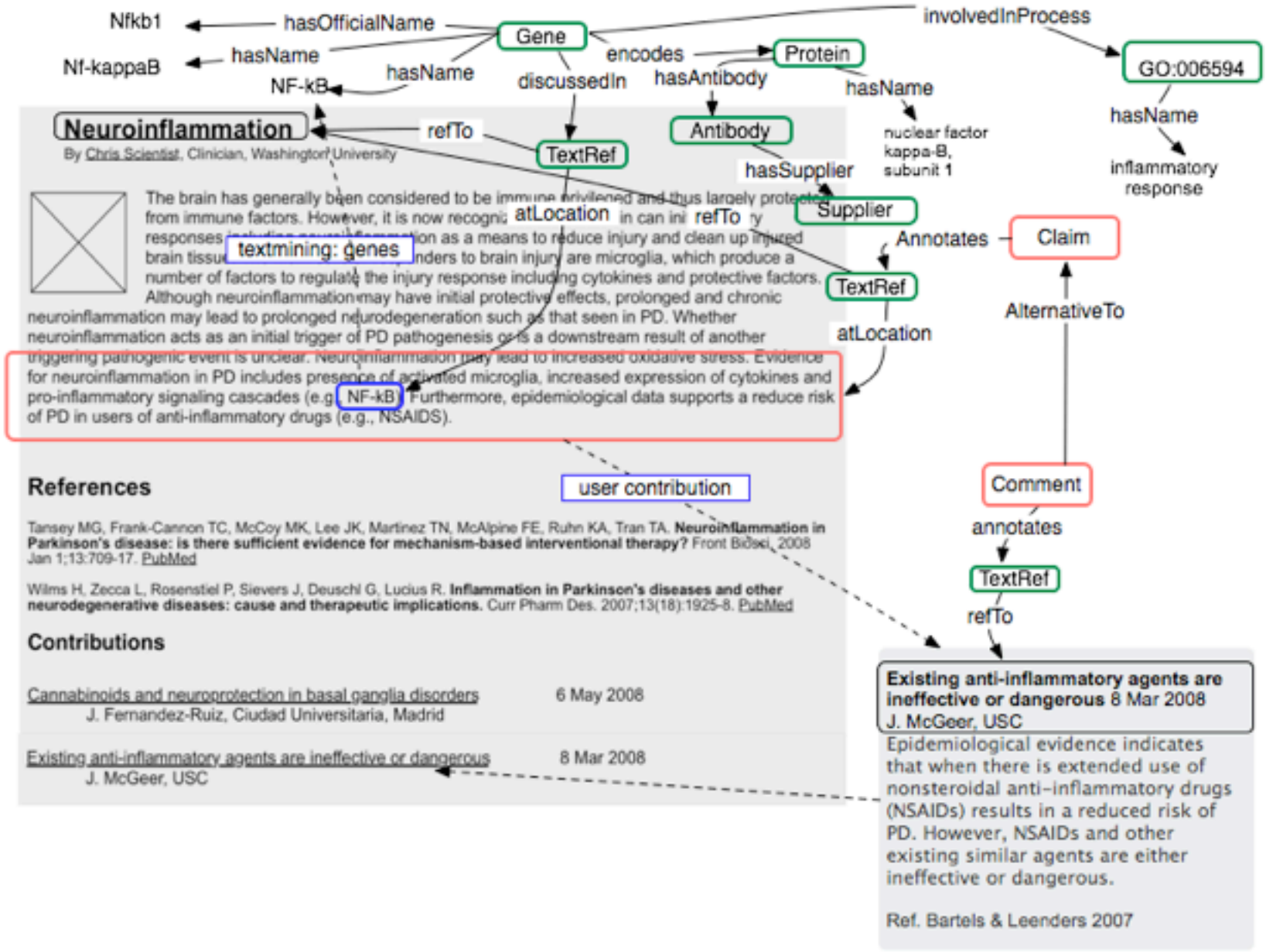
Semi-automatic Annotation



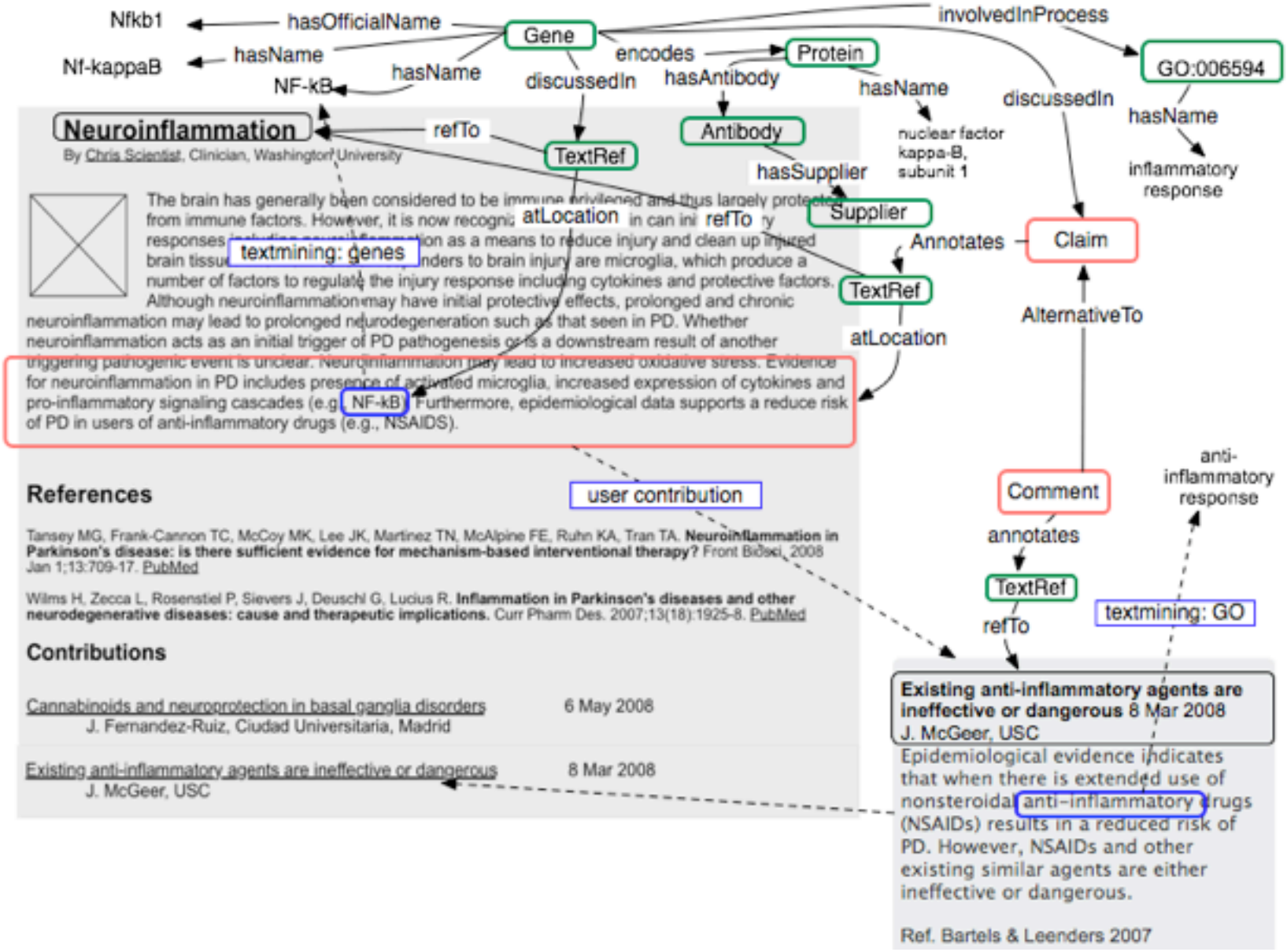
Semi-automatic Annotation



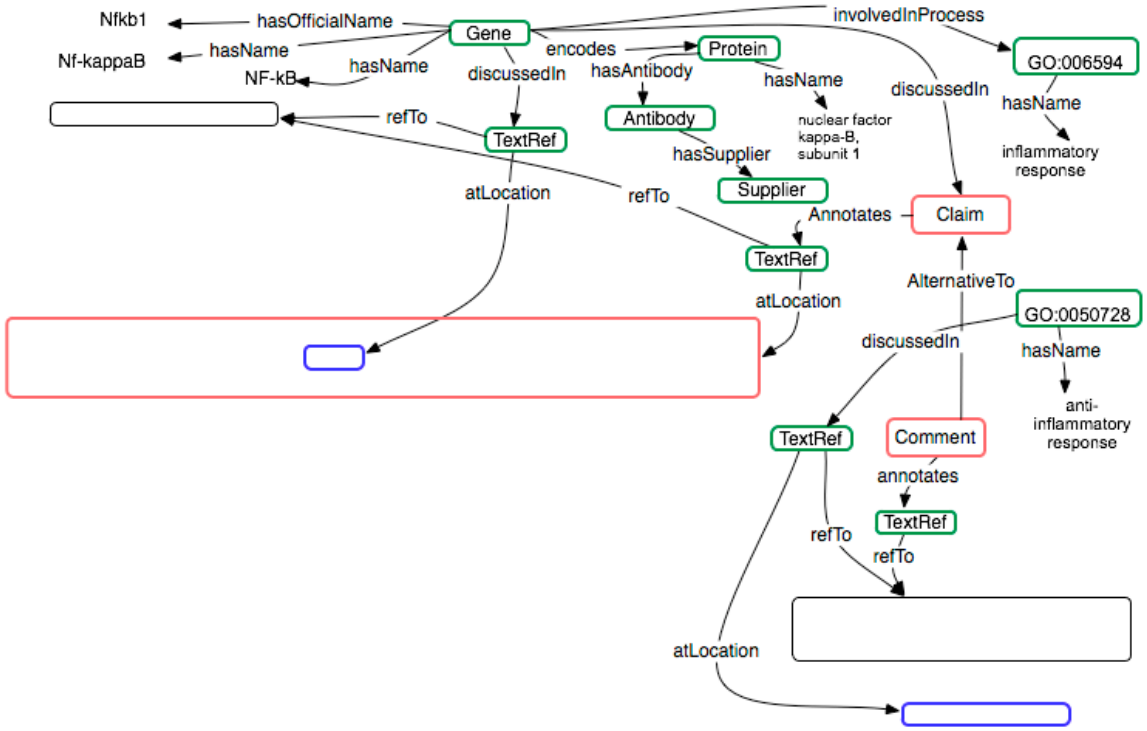
Semi-automatic Annotation



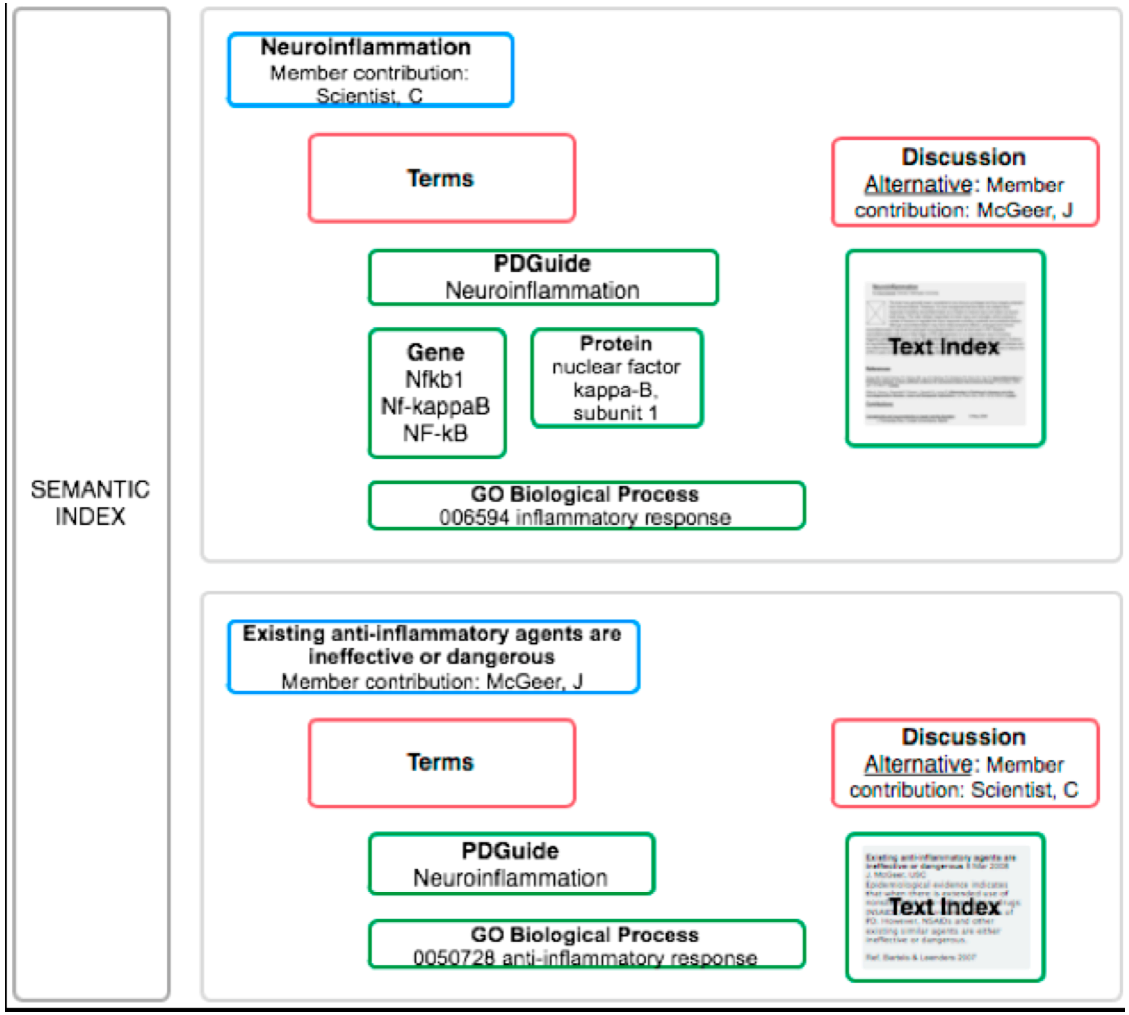
Semi-automatic Annotation



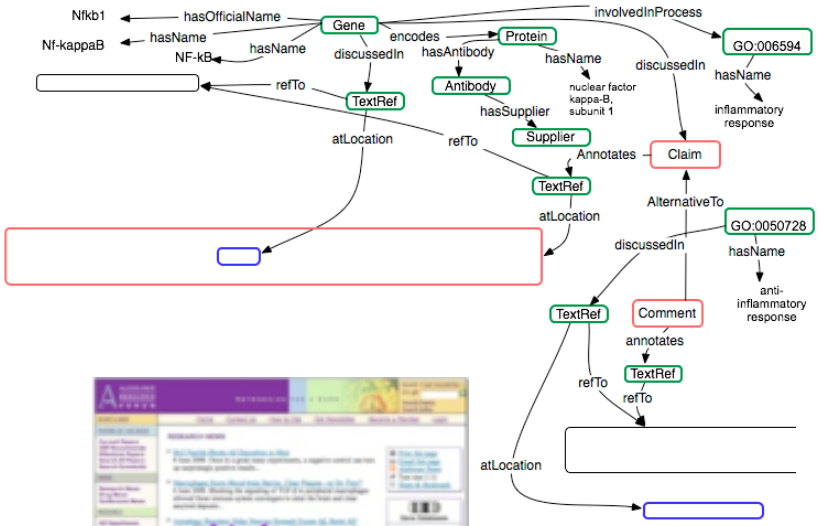
Process the meta-data

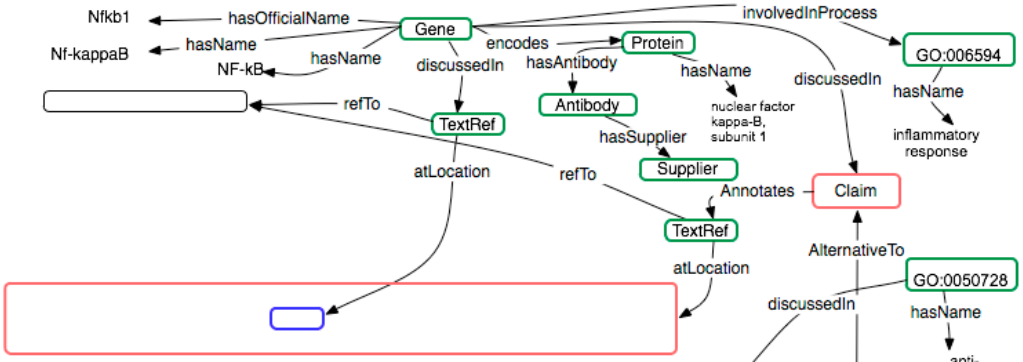


Build Semantic repository



Integrate across communities





Summary

- Reusable framework for building communities
- Easy to adopt
- Leverages linked data
- Uses shared ontologies/vocabularies
- Interoperable with other knowledge-bases

Team

IIC

Software Engineer

- Mark Goetz
- Stéphane Corlosquet
- Tom Green
- Alister Lewis-Bowen
- Louis Weitzman

Project Management

- Sudeshna Das
- Tim Clark

Agaric Design

- Ben Melancon

HSCI

Lisa Girard

Brock Reeve

NeuroCommons

Alan Ruttenberg

Jonathan Rees

Alzforum

June Kinoshita

Elizabeth Wu