User Interfaces to Serve as Inspiration for CoVaSEA

Shiladitya Dutta

January 20, 2018

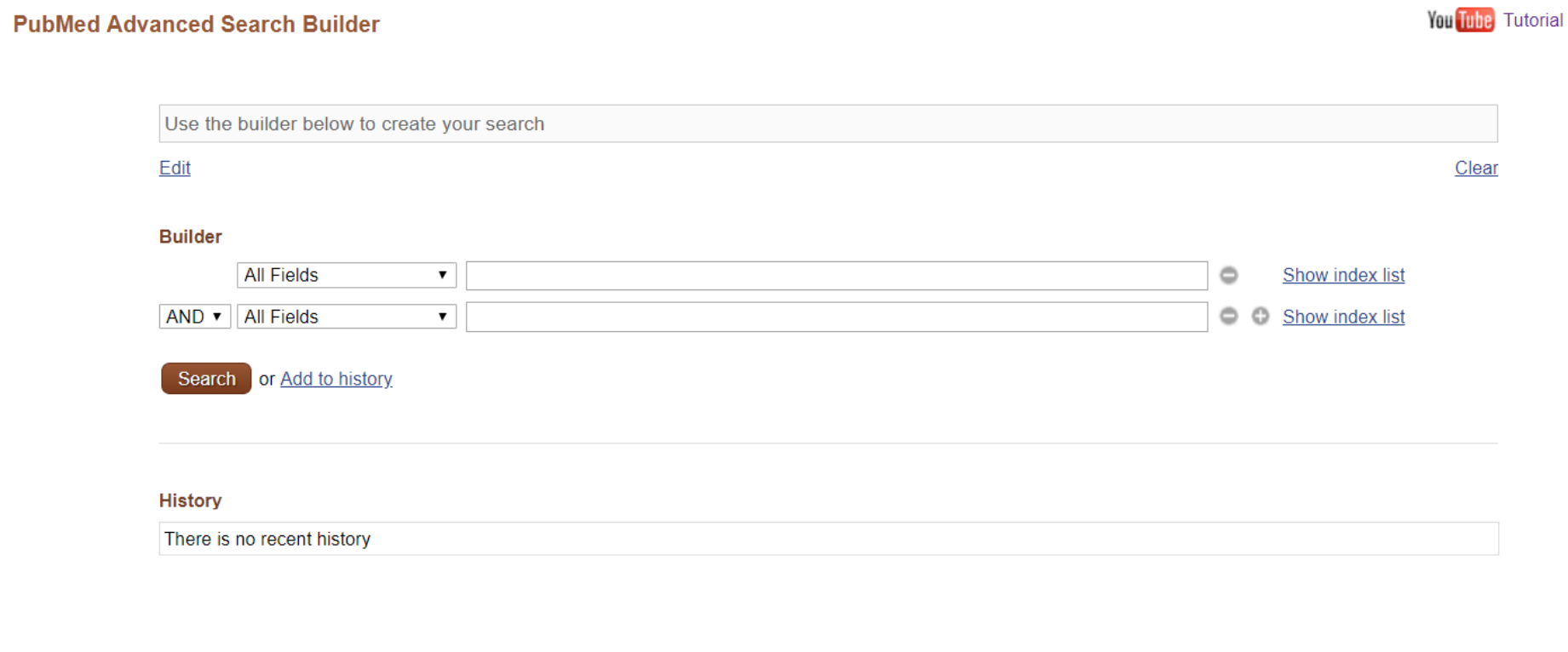
Introduction

- Analyze query builders offered by 4 major research search systems: NCBI PubMed, Elsevier ScienceDirect, Retraction Watch, and Google Scholar
- All of these offer advanced search functionality in some form or another which is designed to provide superior accuracy to that of a general search query
- Analysis consists of the role of the advanced search query and user review of the query
- These will hopefully serve as an inspiration for CoVaSEA's query builder, both for the general and semantic query

NCBI PubMed

- A series of conditions are chained together via AND/OR statements in order to create a search query.
- Also has the option of being inputted directly into a search box. [Ex: (sensory OR language OR motor OR behavior) AND (onset) AND (neurodegenerative OR dementia)]
- This method is relatively intuitive for a user to use once the basic concept is grasped
- Also has a lot more functionality than the conventional method because of the MEDLINE database along with integration with MeSH headings and subheading. (Lowe and Barnett, 1994)
- Overall, this is an extremely proficient system which empowers users of the advanced search functionality greatly.
- System was mirrored by CoVaSEA for the concept-validating system (Bae et. al, 2017).

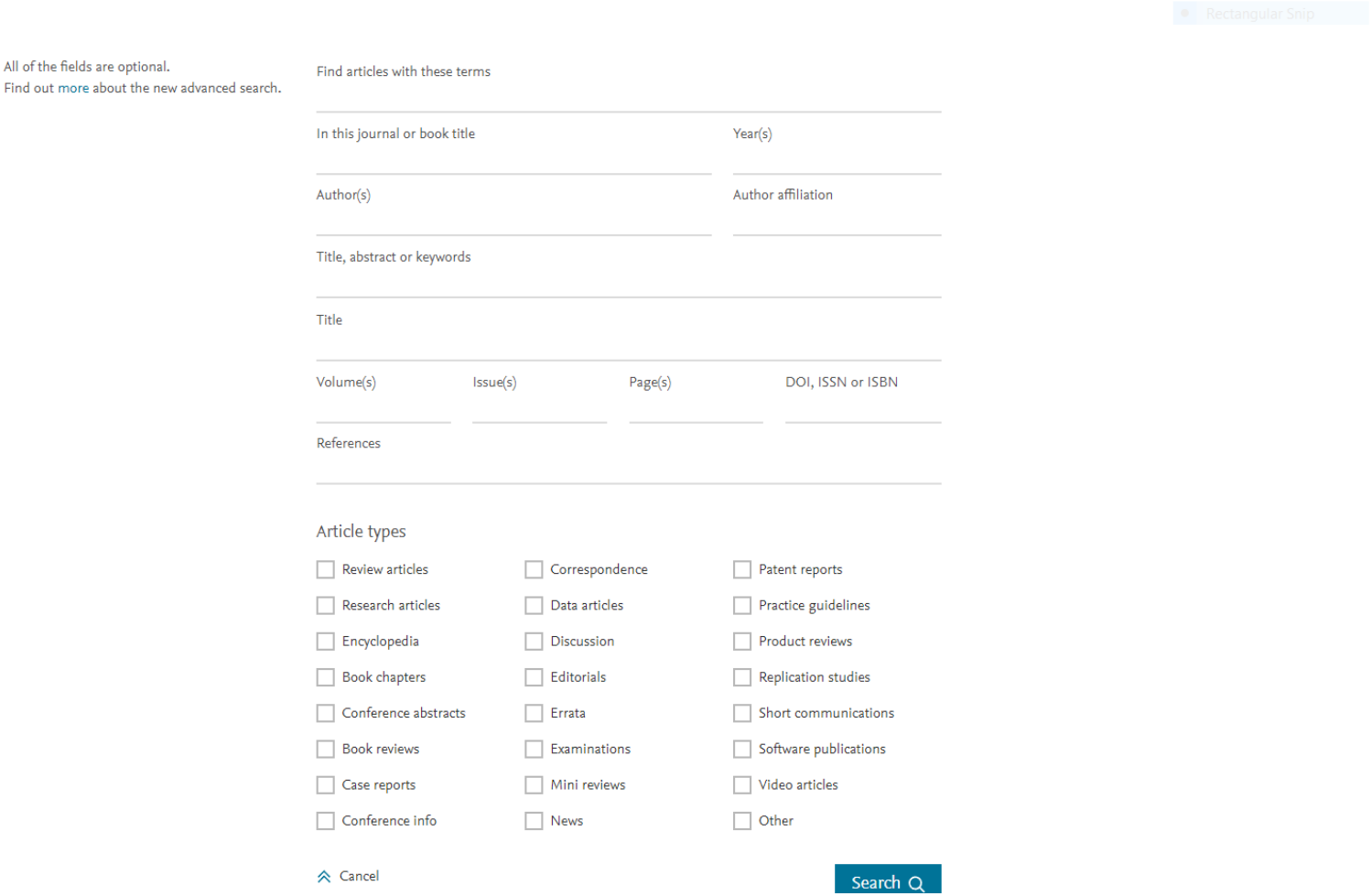
NCBI PubMed Cont.



Elsevier ScienceDirect

- Elsevier ScienceDirect's advanced search works off of a more conventional system of having the option to input a few more conditions
- However, there is one interesting feature of Elsevier's approach. This is the option to specify which type of content the user would like
- The advanced search option fields an exhaustive list of possible content types which can be accessed
- This allows users to specify with high accuracy what general type of content they would like.
- This method may not be exactly transferable to CoVaSEA because of the focus on the categorization of the content rather than the content itself, however it is nevertheless an effective approach.

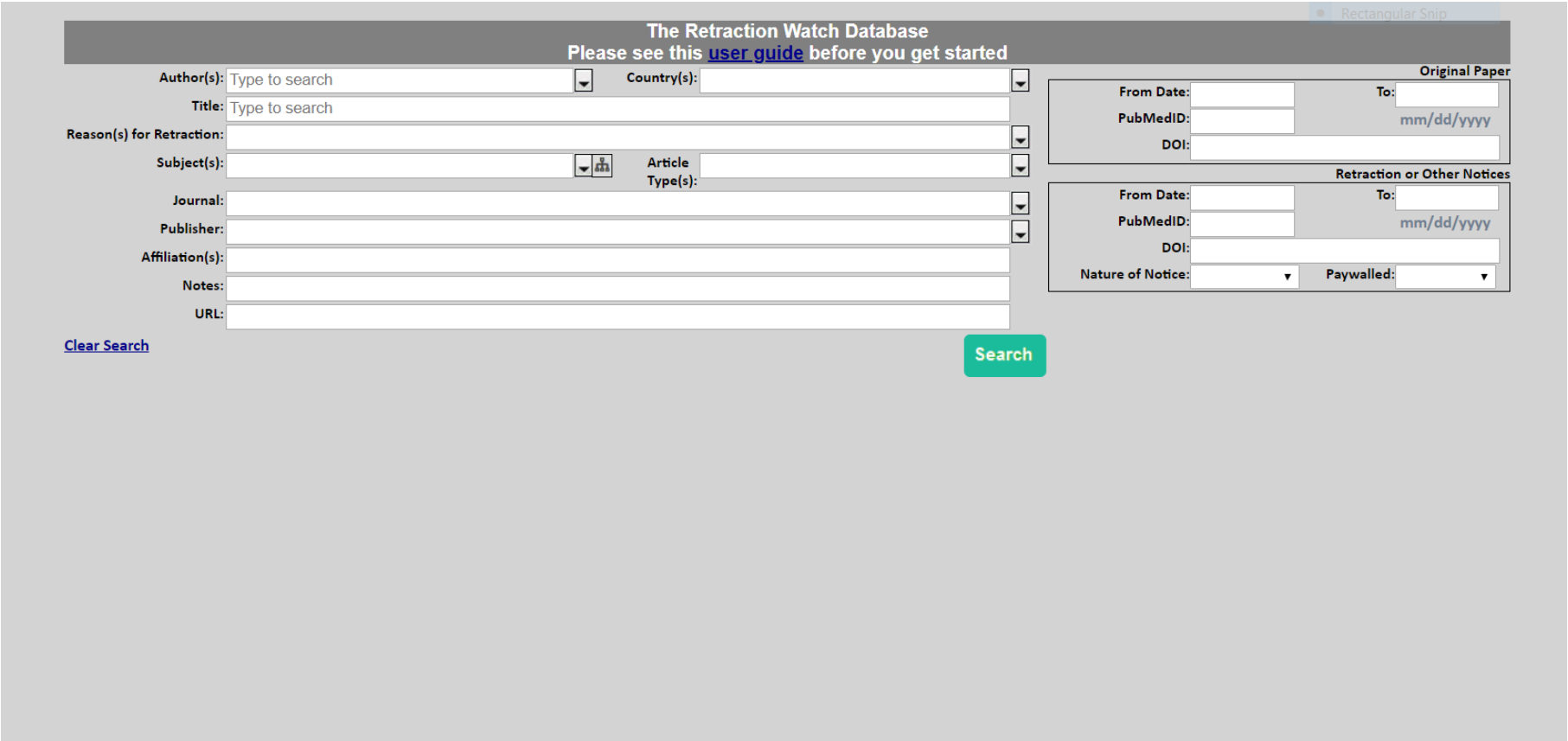
Elsevier ScienceDirect Cont.



Retraction Watch

- Retraction Watch's search system is less focused on searching based on the content of a paper and more focused on narrowing down based on citation metadata.
- Matches with objective of Retraction Watch which is to provide a database of retracted scientific literature.
- Also has support for entering the PubMedID in order to search for redacted papers.
- Should be noted to be a bit daunting at first glance to an average user because there is no option for a simpler search
- The search method is definitely effective, if somewhat lacking on surface level elegance

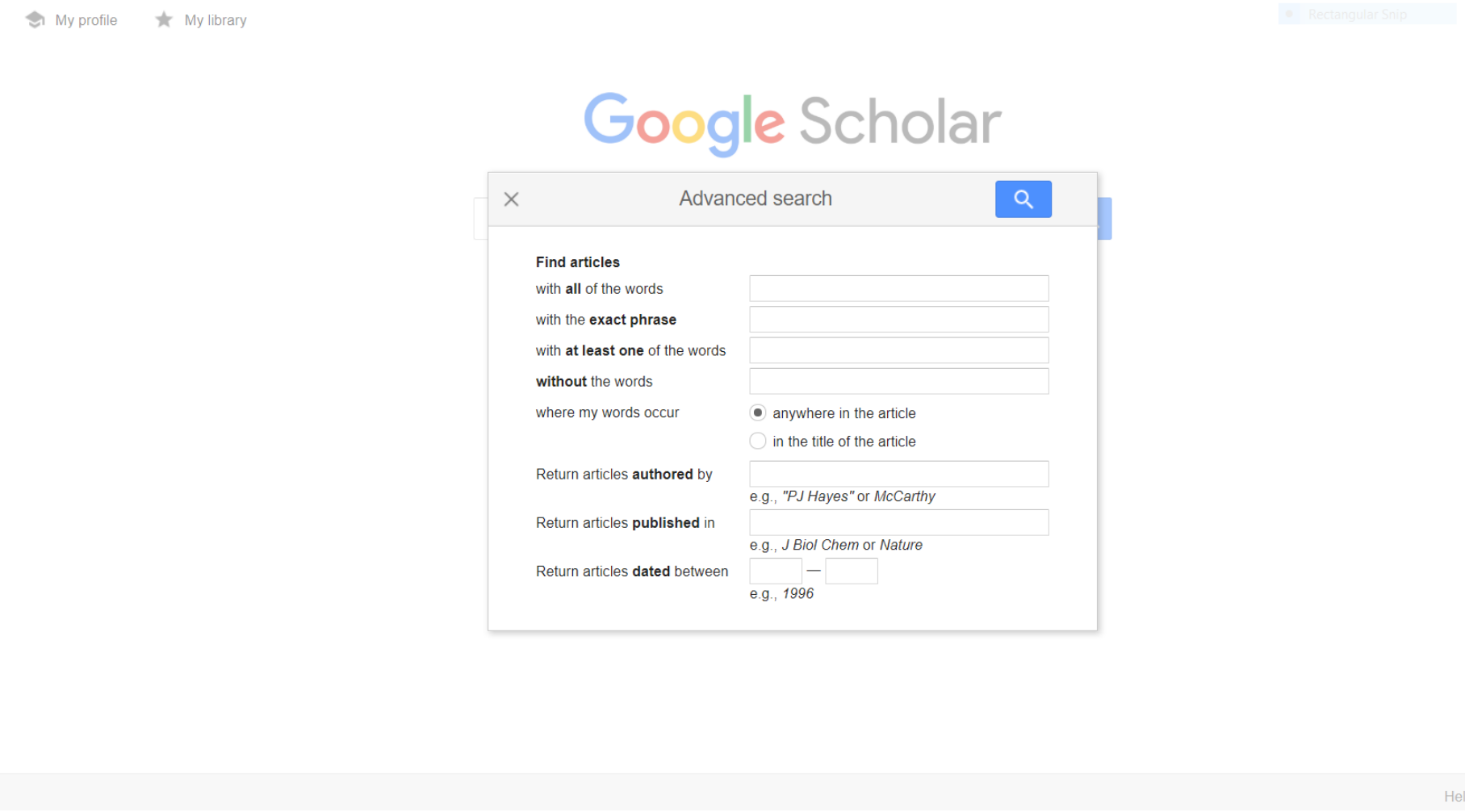
Retraction Watch Cont.



Google Scholar

- Google Scholar is the search system offered by Google
- Google Scholar differs somewhat from the other systems in that it isn't focused purely on scientific literature, but also patents and books.
- This wide breadth of content makes it significantly more difficult for an effective advanced search builder to be made because there are a wide variety of factors to take into account
- Google solves this by opting for a more generalized version of having multiple conditions
- This generalized system comes in the form of a series of options regarding word matching.
- This method is heavily dependent on the lexical analysis side, which makes sense since lexical search is Google's main business.
- However, this won't be transferable to CoVaSEA as this search method isn't necessarily that effective for users due to the lack of ability to narrow down based on citation data or on general topic.

Google Scholar Cont.



Inspiration from Examples

- CoVaSEA ideally should be a combination of Retraction Watch's approach and PubMed's approach
- Retraction Watch provides an excellent variety of conditions to narrow down results with, which should be useful for the web crawler section of CoVaSEA
- PubMed offers inspiration for the SPARQL Query builder since both are inherently condition based.
- Mirroring PubMed's approach of listing conditions should be very effective for a simple to use SPARQL Query builder
- There are also other SPARQL query builders to take inspiration from such as Wikidata's SPARQL Query builder and SPARQL Views (Vrandečić et. al, 2014) (Clark, 2010)
- Most of these visual SPARQL Query builder use a similar condition listing approach

Future Expansions

- One of the main expansions to CoVaSEA would be to integrate the general search query with the SPARQL search query
- This would be accomplished via the translation of the general search query to a SPARQL Search Query
- Research has been done on this topic (Ngonga Ngomo et. al, 2013) (Wang et. al, 2007) (Yahya et. al, 2012)
- Most of these depend upon the parsing of generic search query into a tree for translation which can then have a SPARQL Query extracted from it.
- This would then lead naturally into the final stage of the CoVaSEA user interface: text-to-speech recognition
- The numerous off the shelf solutions which are available would make text-to-speech relatively easy to implement once the natural language to SPARQL Query translation is complete

References


S.-H. Bae, A. G. Craig, C. Taswell, *et al.*, "Expanding nexus diristries of dementia literature with the npds concept-validating search engine agent,",


H. J. Lowe and G. O. Barnett, "Understanding and using the medical subject headings (mesh) vocabulary to perform literature searches," *Jama*, vol. 271, no. 14, pp. 1103–1108, 1994.


D. Vrandečić and M. Krötzsch, "Wikidata: A free collaborative knowledgebase," *Communications of the ACM*, vol. 57, no. 10, pp. 78–85, 2014.


A.-C. Ngonga Ngomo, L. Bühmann, C. Unger, J. Lehmann, and D. Gerber, "Sorry, i don't speak sparql: Translating sparql queries into natural language," in *Proceedings of the 22nd international conference on World Wide Web*, ACM, 2013, pp. 977–988.


C. Wang, M. Xiong, Q. Zhou, and Y. Yu, "Panto: A portable natural language interface to ontologies," in *European Semantic Web Conference*, Springer, 2007, pp. 473–487.


M. Yahya, K. Berberich, S. Elbassuoni, M. Ramanath, V. Tresp, and G. Weikum, "Natural language questions for the web of data," in *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, Association for Computational Linguistics, 2012, pp. 379–390.


L. Clark, "Sparql views: A visual sparql query builder for drupal," in *9th International Semantic Web Conference, ISWC*, 2010.