

Expanding Nexus Diristries of Dementia Literature with the NPDS Concept-Validating Search Agent

Seung-Ho Bae, Adam G. Craig, and Carl Taswell
Brain Health Alliance, Ladera Ranch, CA

Abstract

Even though online databases make it easier than ever to access the biomedical and scientific literature about dementia, accelerating growth in the size of these databases has made it more difficult for humans to gather and analyze manually all articles relevant to any given topic. We document a Nexus-PORTAL-DOORS System (NPDS) Concept-Validating Search Engine Agent that can populate Nexus diristries with concept-validated metadata records for citations of journal articles found in literature databases.

Objectives

Software agents known as crawlers have been developed to automatically retrieve web documents through standard HTTP protocol to index the vast volume of literature on the Internet [1]. In automated information retrieval services built to search for scientific literature, an important objective is to prevent the need for a human user to spend long periods of time conducting manual literature searches. Furthermore, the ability for these software agents to retrieve many articles within a short period of time may help in finding resources that would otherwise not be found in a manual search.

Nexus-PORTAL-DOORS

The Nexus-PORTAL-DOORS System (NPDS) offers an approach by which individuals and organizations can establish independent repositories of semantic and lexical metadata about resources relevant to a problem domain of interest [2]. NPDS specifies a RESTful API for record search and retrieval and a messaging protocol that maintains a flexible but consistent structure for metadata records, enabling exchange among client applications and networks of servers organized according to the Hierarchically Distributed Mobile Metadata architectural style [2]. Nexus diristries are hybrid servers that serve both the semantic descriptions from DOORS directories and the lexical metadata from PORTAL registries.

Concept-Validating Constraints

Concept-validating constraints offer an intuitive way to define a problem domain of interest. To create a validation test, the domain expert selects IRI labels from controlled vocabularies or plain-text word, word-stem or phrase tags and groups them into expressions in conjunctive normal form, e.g. (sensory OR language OR motor OR behavior) AND (neurodegenerative OR dementia). A Nexus record must feature at least one item from each group [3]. A server can use multiple tests with different types of tags or labels, in which case the record need only pass one of the tests to be valid [3].

Table I: Validation pass-rates for CoVaSEA search results.

Diristries	Databases			
	PubMed	DOAJ	Springer	Elsevier
SOLOMON	447/500 (89.4%)	500/500 (100%)	182/500 (36.4%)	421/500 (84.2%)
BRAINWATCH	454/500 (90.8%)	488/500 (97.6%)	176/500 (35.2%)	492/500 (98.4%)

Fig 1: CoVaSEA operations across the Nexus-PORTAL-DOORS System and other databases.

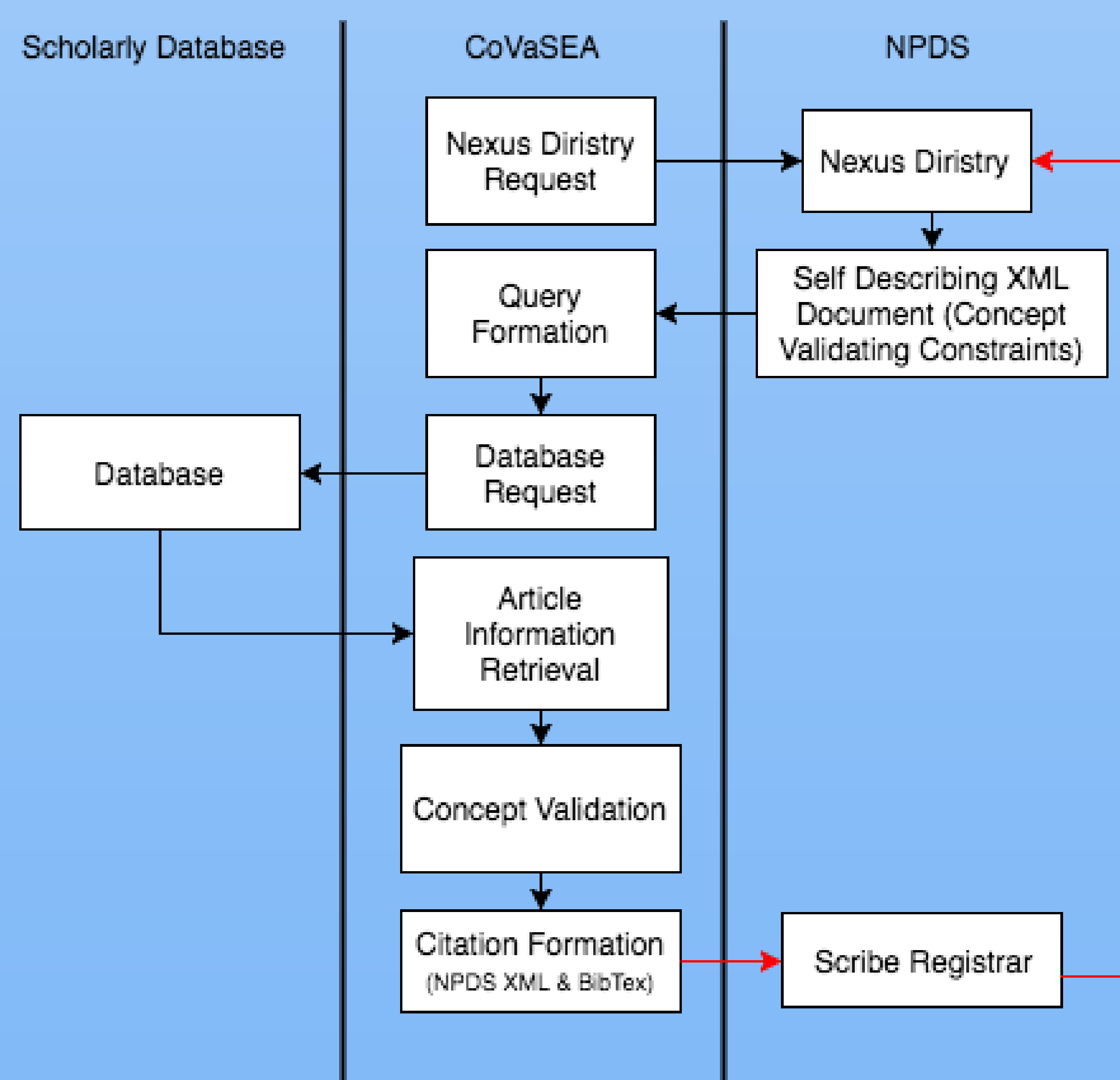
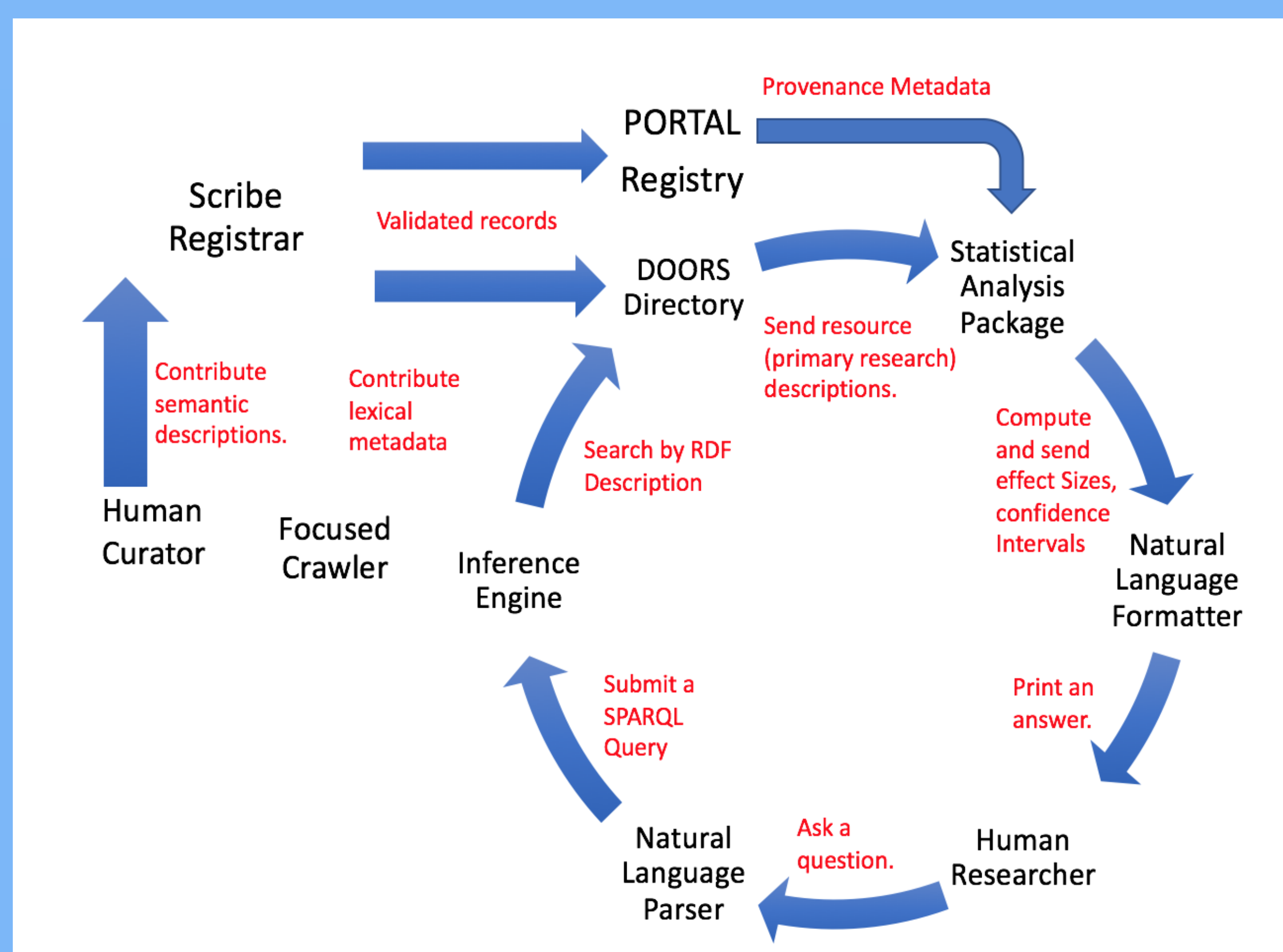


Fig 2: Proposed architecture of automated meta-analysis component built upon the Nexus-PORTAL-DOORS System



CoVaSEA

The NPDS Concept-Validating Search Engine Agent (CoVaSEA) is a JavaScript application that is able to retrieve the self-describing record from a Nexus diristry in order to extract a problem domain's concept-validating constraints. Individual vocabulary terms can also be manually extracted by a human user to vary searches from a specified Nexus diristry, and gather a higher density of relevant results. CoVaSEA uses the domain's key terms to search an online database, retrieve resource metadata for each search result, record it into a Nexus record, and concept-validate each record. BibTeX formatted citations are also able to be created from the article information retrieved by CoVaSEA.

Results

Table I shows the validation pass-rates for search results from four databases, the National Library of Medicine's PubMed, the Directory of Open Access Journals (DOAJ), Springer International Publishing's Springer Link, and Elsevier's Science Direct, against two Nexus diristries: BrainWatch, which covers brain imaging, brain informatics, and neuropsychiatry, and SOLOMON, which focuses on focal-onset neurodegenerative diseases. For each of the four databases, a sample search of 500 articles was run using query terms from the SOLOMON and BrainWatch diristries. The number of relevant articles explained by each Nexus diristry's concept validating constraints was then recorded as a proportion. Each article's title and abstract were considered for concept validation testing. By searching multiple databases and maintaining a consistent standard of relevance, CoVaSEA can fill a Nexus diristry with a high quantity of high-quality records.

Discussion

NPDS metadata records about literature in medicine and science should be curated with additional lexical/semantic markup by human editors. CoVaSEA maximizes the efficiency of knowledge curation by automating the process of finding relevant resources, thereby increasing the amount of time available for human curators to improve the quality of resource metadata. A future component that is able to orchestrate searches and find conclusions to hypotheses based on well-curated metadata can be integrated with CoVaSEA and NPDS in order to produce a search engine built upon an infrastructure system with a hierarchical distribution of metadata and messaging exchange standard that is absent in traditional search engines (Figure 2).

References

1. S. Batsakis, E. Petrakis, and E. Milios. Improving the performance of focused web crawlers. *Data & Knowledge Engineering*, 68(10):1001-1013, 2009.
2. C. Taswell. Doors to the semantic web and grid with a portal for biomedical computing. *IEEE Transactions on Information Technology in Biomedicine*, 12(2):191-204, 2008.
3. C. Taswell. Concept validating methods for maintaining the integrity of problem oriented domains in the portal-doors system, 2010.