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

Seung-Ho Bae[1], Adam G. Craig[1], Carl Taswell[2], *Senior Member IEEE*

***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.**

I. Introduction

Software agents known as focused crawlers have been developed to retrieve online resources relevant to a given topic via standard internet protocols [1]. We report here the development of a novel focused crawler, called the Concept-Validating Search Engine Agent, that can access the vast volume of biomedical and scientific literature available online and that can interoperate with NPDS Nexus diristries [2]. NPDS offers an approach by which independent repositories of semantic and lexical metadata can be established for resources relevant to a problem-oriented domain of interest [3]. NPDS specifies a RESTful API [2] 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 HDMM architectural style [3].

II. Methods

Concept-validating constraints offer an intuitive way to define a topical field of interest [4]. To create validation tests, the domain expert selects IRI labels from controlled vocabularies and/or plain-text tags as word-stems, words or phrases, and groups them into expressions in conjunctive normal form, e.g. (sensory OR language OR motor OR behavior) AND (onset) AND (neurodegenerative OR dementia) for the SOLOMON diristry. The NPDS Concept-Validating Search Engine Agent, called CoVaSEA, is a JavaScript application that retrieves the self-describing record from a Nexus diristry in order to extract the concept-validating constraints for its problem-oriented domain. CoVaSEA uses the key terms in the constraints to search an online database, retrieves resource metadata for each reference citation, formats it as a Nexus metadata record, and validates the concepts to test the metadata record for possible inclusion in the Nexus diristry.

*This work was done by students at the Brain Health Alliance Virtual Institute (BHAVI), a 501-c-3 not-for-profit organization.

[1]Seung-Ho Bae (sbae@bhavi.us) and Adam G. Craig (acraig@bhavi.us) are students at Brain Health Alliance Virtual Institute.

[2]Carl Taswell (ctaswell@brainhealthalliance.org) serves as director at BHAVI.

III. Results

Table I summarizes the concept-validation success rates for search results from three databases, the National Library of Medicine's PubMed, the Directory of Open Access Journals (DOAJ), and Springer Publishing's SpringerLink, when tested for concept validation against two Nexus diristries, BrainWatch for brain imaging and neuro-informatics, and SOLOMON for dementias and neurodegenerative diseases. By searching multiple literature databases and maintaining a consistent standard of relevance, CoVaSEA can fill a Nexus diristry with high-quality metadata records that have passed the concept-validation tests for inclusion in the diristry.

TABLE I
Concept-validation rates for CoVaSEA

Diristries	Literature Databases		
	PubMed	DOAJ	SpringerLink
SOLOMON	46/50 (92%)	50/50 (100%)	16/50 (32%)
BrainWatch	50/50 (100%)	50/50 (100%)	26/50 (52%)

IV. Discussion

To build effective knowledge engineering applications, it is not sufficient merely to collect resources. Human editors should also curate metadata records about these resources by contributing additional metadata. CoVaSEA can help maximize the efficiency of human curators for NPDS metadata records by automating the process of finding and collecting relevant resources, thereby enabling these expert curators to spend more time editing the additional revisions and commentary in the metadata records that only humans can provide. We plan to integrate CoVaSEA with NPDS and other future software components to build a search engine capable of orchestrating searches for biomedical hypotheses about dementia based on expert-curated metadata in repositories that are compliant with NPDS and its open messaging exchange standard that is absent in traditional search engines.

References

[1] S. Batsakis, E. Petrakis, and E. Milios, "Improving the performance of focused web crawlers," *Data & Knowledge Engineering*, vol. 68, no. 10, pp. 1001–1013, Oct. 2009.

[2] A. Craig, S.-H. Bae, T. Veeramacheneni, K. Taswell, and C. Taswell, "Web service APIs for Scribe registrars, Nexus diristries, PORTAL registries and DOORS directories in the NPD System," in *SWAT4LS 2016 Amsterdam NL*, 2016.

[3] C. Taswell, "A distributed infrastructure for metadata about metadata: The HDMM architectural style and PORTAL-DOORS system," *Future Internet*, vol. 2, no. 2, pp. 156–189, June 2010.

[4] ——, "Concept validating methods for maintaining the integrity of problem oriented domains in the PORTAL-DOORS System," in *IDAMAP 2010 Workshop Washington DC*, 2010, pp. 78–79.