

AlzPharm: A Light-Weight RDF Warehouse for Integrating Neurodegenerative Data

Kei-Hoi Cheung^{1,2,3,4}, Hugo Y.K. Lam⁵, Luis Marengo^{1,2}, Tim Clark^{8,9}, Yong Gao⁹, June Kinoshita¹⁰
Gordon Shepherd⁶, Elizabeth Wu¹⁰, Gwen Wong¹⁰, Nian Liu^{1,2}
Chiquito Crasto^{1,2}, Thomas Morse⁶, Susie Stephens¹¹, and Perry Miller^{1,2,7}

¹Center for Medical Informatics, ²Anesthesiology, ³Genetics, ⁴Computer Science

⁵Program in Computational Biology and Bioinformatics, ⁶Neurobiology

⁷Molecular, Cellular and Developmental Biology, Yale University, New Haven, CT, USA

⁸Harvard University, Initiative in Innovative Computing, Cambridge, MA, USA

⁹Massachusetts General Hospital, Boston, MA, USA, ¹⁰Alzheimer Research Forum, ¹¹Oracle

Research progress in Alzheimer's Disease and other neurodegenerative diseases is entering a critical stage in which it is essential to carry out scientific investigation at multiple levels (including genetic, molecular, and cellular levels) in order to gain new and deeper insights into the physiopathology underlying these complex diseases. To take full advantage of these experimental initiatives, it is important to create databases that enhance the ability of different laboratories to access, exchange, and integrate such diverse data. It is also important to integrate such data from fundamental research with human genetic and clinical data to enhance the effort to understand, prevent, and treat these disorders.

We have built a light-weight data warehouse using the Oracle RDF Data Model (http://www.oracle.com/technology/tech/semantic_technologies), acting as a semantic glue between BrainPharm (<http://senselab.med.yale.edu/senselab/BrainPharm/alzData.asp>) of SenseLab (<http://senselab.med.yale.edu/senselab/>) and AlzForum (<http://www.alzforum.org/>) through SWAN (<http://swan.mindinformatics.org/>). Instead of translating all of the data in these databases into a central data warehouse, we have converted subsets of data into RDF, allowing us to evaluate this approach. This "thin" data warehouse approach is potentially more scalable and maintainable in terms of integrating a large number of databases. For retrieval of more detailed information, we rely on the native web-based user interface provided by each database. As a demonstration, we have developed a Web-based application called "AlzPharm" (<http://ontoweb.med.yale.edu/AlzPharm>) that allows to integrate drug-related information from BrainPharm with publication data stored in AlzForum. Figure 1 shows the demo application components including data conversion and loading into the Oracle (RDF) database. The figure also shows the Web query interface that allows the user to pose an integrated query across BrainPharm and SWAN datasets. The query results include links to BrainPharm as well as links to AlzForum for detailed drug/publication information. Our Semantic Web approach extends the data warehouse solution by providing a richer semantic representation and inferencing capability. It can also be adapted to other integration solutions involving many different types of neuroscience information.

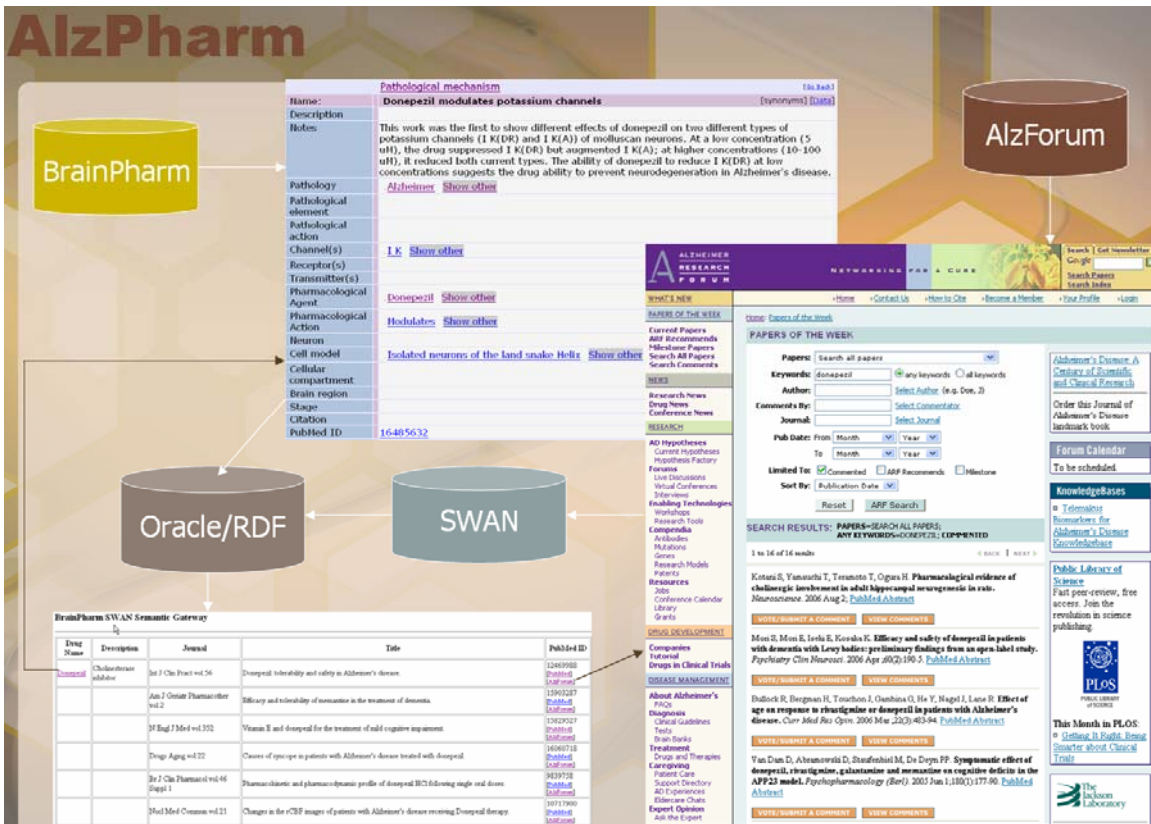


Figure 1. The system components of AlzPharm.