<p>WordNet® is a large lexical database of English. Nouns, verbs, adjectives and adverbs are grouped into sets of cognitive synonyms (synsets), each expressing a distinct concept. Synsets are interlinked by means of conceptual-semantic and lexical relations. Wordnet® is a trademark of Princeton University. See Princeton University “About WordNet”; WordNet®. Princeton University. 2010. <a href=*"http://wordnet.princeton.edu"*></a></p>

<p>This example application is an adaptation of Wordnet® for Apache HBase™ using CloudGraph™, a suite of Java™ standards-based data-graph mapping and ad hoc query services for big-table sparse, columnar "cloud" databases.</p>

<p>For more information on CloudGraph™, see <a href=*"http://cloudgraph.org"*>http://cloudgraph.org</a>

<p>This adaptation of Wordnet® for HBase™ was accomplished in 5 basic steps using CloudGraph™ and related tools.</p>

<p>1.) Model Creation. First the Wordnet® relational MySql® database schema was automatically reverse engineered and converted to UML for CloudGraph™ using Plasma and PlasmaSDO™ Relational Database (RDB) provisioning tools. Models can be easily hand written as well.</p>

<p>2.) Code Generation. Using PlasmaSDO™ both a Service Data Objects (SDO) persistence layer and Domain Specific Language (DSL) query layer were generated from the Wordnet® UML model. These generated packages are used in the application to query and update the HBase™ data store using the CloudGraph™ HBase™ Service. Javadocs for the generated code can be browsed in the Javadocs section of this application.</p>

<p>3.) HBase™ Table Mapping. For this application, all Wordnet® RDB tables and views were collapsed into a single target HBase™ table. Any number of HBase™ tables could have been used but Wordnet® is often used as part of larger applications, so to economize on tables (typical rule-of-thumb is no more than 10 tables per application) a single table was used. Using the CloudGraph™ configuration facility, all Wordnet® data graphs were mapped to the HBase™ data-store table(s) with specific model root types and composite row key definitions.</p>

<p>4.) Data Migration. Using the generated persistence and DSL query packages, a custom data migration stand-alone application was written to read from the Wordnet® relational database using CloudGraph™ RDB service and incrementally write to the HBase™ data store using the CloudGraph™ HBase™ Service. </p>

<p>5.) Web Application Creation. Finally with an HBase™ data store populated with all the rich semantic and lexical data from Wordnet®, 2 services and a JSF based web application were written using the new AJAX and JQuery based PrimeFaces JSF component set. The first service is called from the auto-complete field which triggers CloudGraph™ HBase™ to use the HBase™ <a href=*"http://hbase.apache.org/apidocs/org/apache/hadoop/hbase/filter/FuzzyRowFilter.html"*>fuzzy row filter</a> API. Then using the selected word, the 'find' button calls a second service which uses a large DSL query spanning all semantic and lexical relations for the given word, as well as several other entities such as example sentences. The DSL query spans more than 7 nodes from the source word to its related target words looking (very roughly) like Words->Senses->Synsets->Semlinks/Lexlinks->Synsets->Senses->Words. Larger word relation graphs typically contain over 100 nodes and are assembled in the middle tier in an average of 200 milliseconds from a 3 node CentOs-Linux HBase cluster.</p>

<p>WordNet® is a registered trademark of Princeton University. HBase™ is a trademark of Apache Software Foundation. MySql® is a registered trademark of Oracle Corporation. CloudGraph™ is a trademark of TerraMeta Software Inc. </p>