SHIRASE: Automated Construction and Analysis of Semantic Graphs Derived From Open-Web Resources for a Hybrid Lexical-Semantic Search Engine

Shiladitya Dutta
Foothill High School, Pleasanton, CA
Teacher: Shannon Sos, Foothill High School

The Semantic Web (Web 3.0) is the next step of the world wide web. Web 3.0 represents a significant improvement over the lexically-based Web 2.0 by employing a data exchange format, semantic graphs, that allows software to understand the data that it's processing, empowering users and software applications alike. However, full implementations of semantic search haven't hit the mainstream due to the high upfront investment required for the creation of semantic graphs and the lack of conducive search methods. In seeking to ameliorate these problems, the Semantic-lexical Hybrid Information Retrieval And Search Engine (SHIRASE) has been developed as a solution targeted towards researchers. SHIRASE syphons articles from literature databases based on user searches, converts the articles to semantic records, and analyzes these records via a SPARQL-based search engine. Because the system remembers all previously converted records, SHIRASE can dynamically accumulate a semantic database powered by user searches. Proverbially hitting two birds with one stone, this end-to-end article search engine allows users to access the breadth of content in the lexical format with the versatility of a semantic search all in a straightforward software interface, thus acting as the bridge which connects the lexical and semantic web.