



Semantic Search and Conceptual Indexing

What's **Squiggle**?

Squiggle is a framework that supports the building of a domain-aware semantic search engine. **Squiggle** represents an abstraction for people who want to build a search engine in a particular domain and do not want to deal with low-level indexing and storing processes.

Squiggle seamlessly combines the speed of syntactic search tools with improved recall and precision. This is because **Squiggle** is able to trace any alternative/multilingual/misspelled labels back to the corresponding concepts, i.e. **Squiggle** can identify and recognize meanings.

The **Squiggle** framework is domain independent and can thus be instantiated with and adapted to any domain specific context and ontology. Among the constituents of **Squiggle**, **Sesame** is used as the semantic engine that queries the knowledge base, described in RDF with regard to the **SKOS model**, whereas the syntactic search engine **Lucene** is used, among other things, to quickly perform text searches in literals, which is something that semantic search tools typically cannot do well. Therefore the **Squiggle** architecture lends itself well both to overcome the limitations of purely syntactic approaches and to improve the performance of semantic engines.

Squiggle is not a search engine itself, but it allows users to customize their own engine on the basis of a particular domain knowledge. **Squiggle** is designed to provide both syntactic and semantic indexing and searching primitives.

To prove our approach we built some test-applications on the top of our **Squiggle** framework. We briefly present the most significant test-implementations. Readers that wish to try semantic-searching with our demonstrative applications can do so just following the hyperlinks.



Squiggle Ski Engine

CEFRIEL is Official Supplier of Applied Academic Research of **Torino 2006 Olympic Winter Games**. We have the opportunity to demonstrate **Squiggle** in the context of the CEFRIEL's activities related to the Winter Olympic Games.

A simple way to understand the power of Squiggle is searching for "libera" which is the Italian word for the Alpine Ski "downhill" discipline. Only 33 results are returned, but if you click on the "downhill" link in the "did you mean..." box, you got 515 hits and the explanation of the results.

We built the Domain Knowledge partially by hand, and partially by collecting information on the **FIS-Ski web site**. By hand we developed a small multilingual taxonomy of the disciplines in the sectors of alpine ski, cross country ski and snowboarding. For example we have the concept of "downhill" with labels in Italian ("discesa libera"), German ("Abfahrt"), Swedish ("Störtlopp"), English ("downhill") and so on. From FIS-Ski web site we collected: all the athletes that got a podium in FIS World Ski Championships, FIS World Cup, and Olympic Winter Games; all the event in the last three years and the relationships with the nations that hosted them, the top three athletes of the event and the type of event (e.g., downhill, slalom, giant slalom and combined).

Instructions: try **Squiggle Ski** and write the name of your preferred athlete and his/her discipline!

(please, notice that you'll not find *all* the athlete you know, but only those that got a podium in the Winter Olympic Games and FIS Worlds Championships ...)



Squiggle Music Engine

Squiggle Music is a music search engine on the Web that allows its users to retrieve (information about) songs by keyword searching. Its searching

capabilities include semantic features allowing term disambiguation and query expansion.

A Web crawler searches for music on the Web and stores in an internal database information concerning each file (e.g., file name and URL) as well as meta-tags that might be included in the file descriptors (e.g., title, author, etc.). In total, the archive currently contains information concerning nearly half a million songs.

Two freely available meta-databases developed and maintained by web communities are used to compose Squiggle Music's Domain Knowledge: **MusicBrainz** (from which we took information concerning names of music bands and titles of tracks as well as associations between different bands/artists) and **MusicMoz** (from which we took also a taxonomy of musical styles and associations between bands/artists and styles).

On the user's side, Squiggle Music presents to the user a list of songs as the result of a syntactic search. The answer page also contains a **disambiguation box** that shows possible meanings matching the search, obtained by semantic disambiguation. Meanings in this context may correspond to bands/artists, tracks, or musical styles. If the user selects one of the suggested meanings, Squiggle Music retrieves all songs that correspond to it and all other meanings that have a semantic relationship with it.

For example, if the user query is "RHCP", the system suggests "Red Hot Chili Peppers" as one of the possible meanings ("RHCP" being a known acronym); selecting the proposed meaning results in better recall, since only few songs are likely to match the search string syntactically.

If the user asks for "Rock", Squiggle Music identifies the "Rock" music style; therefore, the search results both in better recall and precision, since all songs by rock artists are retrieved and included in the result set. The user may also indicate whether the search should be extended by looking at super- or sub-genres of a given style, in case the number of retrieved document should be under a certain threshold: in this case, Squiggle Music also includes in the result songs that belong to styles that are immediately above or below in the taxonomy, i.e. those styles that are related through a **skos:broader** or a **skos:narrower** relation with the identified meaning.

Moreover, once the user is satisfied with the result set, Squiggle Music offers the facility to build a playlist with the retrieved songs.

Instructions: try **Squiggle Music** and submit your musical query!
(please, notice that you'll not find *all* the music artists you know...)

People

Squiggle is conceived, designed and implemented by: **Emanuele Della Valle**, **Irene Celino**, **Dario Cerizza** and Davide Martinenghi