# X-Link: A Linked Data-based Named Entity Extraction tool

Overview

Named Entity Extraction (NEE) is the process of identifying entities in texts and, very commonly, linking them to related (Web) resources. Although this task is useful in several applications, most NEE solutions lack an open and easy configuration interface which is very important for building domain-specific applications. X-Link is a fully configurable Linked Data-based Named Entity Extraction tool which allows the user/developer to easily define the categories of entities that are interesting for the application at hand by exploiting one or more online Semantic Knowledge Bases (Linked Data repositories). The user is also able to update a category and specify how to semantically link and enrich the identified entities. This enhanced configurability allows X-Link to be lightly configured for different contexts, for building domain-specific applications (e.g. for identifying drugs in a medical search system, for annotating and exploring fish species in a marine-related web page, etc.).
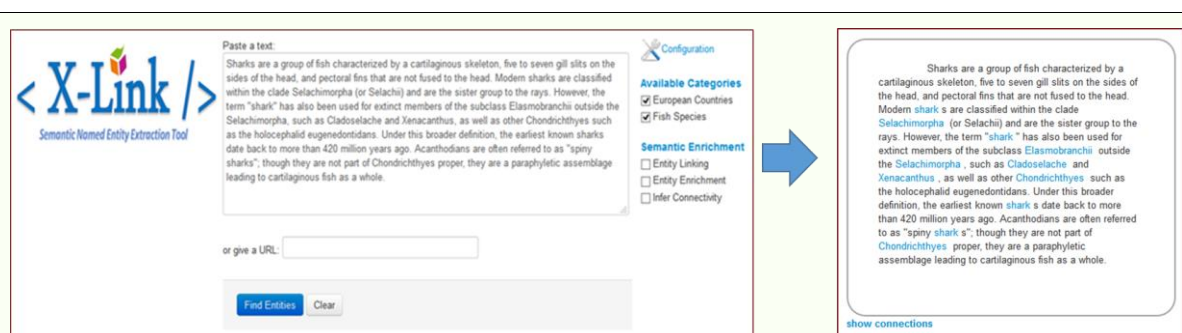

Identifying, extracting and linking the entities of a text using X-Link web application

Target Applications

Named Entity Extraction tools are useful in many applications e.g. for question answering, annotating documents, post-processing of search results, etc.

X-Link is fully configurable in terms of the supported categories of entities, the underlying Knowledge Bases and the way the system queries the Knowledge Bases, which makes appropriate for a wide range of domain-specific applications.
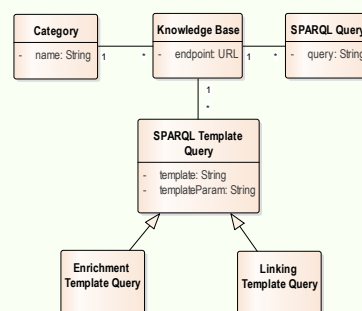
The generic model for configuring X-Link

Description

X-Link is based on Gate ANNIE tool and supports both gazetteers (lists of names) and natural language processing functions. Gate ANNIE is a ready-made information extraction system which contains several components (e.g. Tokenizer, Gazetteer, Sentence Splitter, Orthographic co-reference, etc.). For X-Link, we have extended it, in order to be able to create a new supported category and update an existing one (using gazetteers) by exploiting the Linked Data.

X-Link supports the analysis of plain text files, HTML pages, Microsoft Word and Powerpoint files (.doc, .docx, .ppt and .pptx), PDF files, and XML-based files (e.g. XML and RDF files). X-Link starts by reading an initial configuration which is stored in a properties file. It also implements functions that allow the user/developer to configure the system, e.g. through an administrator API. Specifically, the following functions are currently supported:

- Add a new category, using one or more lists of entities, one or more instances resource classes or one or more instances SPARQL queries. The user can also update an existing category or totally remove one.
- Change the displayed name of an existing category (i.e rename).
- Configure/update the underlying Knowledge Bases
- Configure the SPARQL queries to be sent to the underlying Knowledge Bases for linking the identified entities. Furthermore, the SPARQL queries that are needed for enriching the identified entities can be set.
- Set/change the active categories, i.e. the categories for which X-Link identified entities.

X-Link in iMarine (through X-Search)

X-Link in PerFedPat (through X-Search)

Additional Information

Currently X-Link is used in two different contexts: in the **Marine** domain (in the context of the iMarine project, FP7-283644) and in **Patent Search** (in the context of the PerFedPat project, FP7-275522).

For more information please visit
http://www.ics.forth.gr/isl/X-Link/

X-Link website

Contact details:	**Yannis Tzitzikas**
	tzitzik@ics.forth.gr
	www.ics.forth.gr/isl