

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

Overview

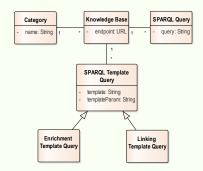
Named Entity Extraction (NEE) is the process of identifying entities in texts and, very commonly, linking them to related (Web) resources. Although this task is useful in several applications, most NEE solutions lack an open and easy configuration interface which is very important for building domain-specific 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.).

show connections	< X-Link /> Semantic Named Entity Estraction Tool	Paste a text: Sharks are a group of fish characterized by a catilaginous skeleton, five to seven gill sits on the sides of the haad, and pectoral fins that are not have to the haad. Modern sharks are classified within the clade StateIntomph or (StateChi) and are the size (sits group to the rays). Howeve, the term "shark" has also been used for extinct members of the subclass Elasmotranchi oxitish the StateChine that a Clade StateChine and are state as well as other Chondrichthyes such as the holocophald eugeneotridans. Under this broader definition, the earliest known thinks date back timore than 420 million years gao. Acanthodians are dens referend to as "spiny sharks", though they are not part of Chondrichthyes proper, they are a puraphyletic assemblage leading to cartilaginous fish as a whole.	Configuration Available Categories Comparison Fish Species Semantic Enrichment Control Linking Control Linking Differ Connectivity		Sharks are a group of 5sh characterized by a cartilaginous skeleton, five to seven gill sits on the sides of the head, and pectoral fins that are not fused to the head. Modern sharks are classified within the clade Selachimopyla (or Selachim) and are the sister group to the rays. However, the term "shark" has also been used for exinct members of the subclass Classrobranchi outside the Selachimorpha, such as Cladoselache and Xenacanthus, as well as other Chondrichthyses such as the holocephaild eugenedontidans. Under this broader definition, the arefirst known shark's date back to more than 420 million years ago. Acanthodans are often referred to as "sign yearks", though what's 'date back to more than 420 million years ago. Acanthodans are often referred to as "sign yearks", though what's 'date back to more than 420 million years ago. Acanthodans are often referred to as "sign yearks", though they are not part of Chondrichthyse proper, they are a paraphyletic assemblage leading to cartilaginous fish as a whole.	
------------------	--	--	--	--	---	--

Target Applications

Named Entity Extraction tools are useful in many applications e.g. for question answering, annotating documents, postprocessing 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.



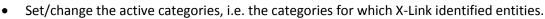
The generic model for configuring X-Link

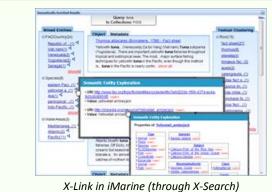
Description

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.







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 <u>http://www.ics.forth.gr/isl/X-Link/</u>



X-Link website

Contact details:

Yannis Tzitzikas tzitzik@ics.forth.gr www.ics.forth.gr/isl

FORTH-ICS