EIT Digital – Industrial PhD position proposal

PhD thesis information

PhD Thesis – Title	**City Enabler for Digital Urban Services (I)**
PhD Thesis – Short summary	This thesis will leverage the City Enabler concept, namely an IT solution that offers a collaborative plaza allowing all relevant private/public stakeholders to work together, both for publishing data and for exploiting such data in order to create novel urban services. In particular, the thesis will investigate innovative approaches, techniques and tools for data and service management, specifically for the "intelligent" retrieval and presentation of information distilled from PA (Public Administration) documents. The thesis will lead to the development of an information retrieval tool, allowing flexible, multi-layered searching of document content and ranking of the search output according to stakeholders' preferences. The results will be evaluated in real-world pilots in the scope of research and industrial projects in the Autonomous Province of Trento.
Rationale/challenge	Nowadays, urban data is closed within silos and as a result it is extremely hard for urban service providers to get access to available and valuable knowledge. There is a strong need for novel approaches that help to open up, analyze and exploit these scattered urban data, covering a wide range of information from different sources (e.g. open, sensor, free, closed, linked data). As for the public administration domain, recent documents are often available in digital format, but the way information can be retrieved and displayed is usually not flexible, and users searching for information and navigating through the archives spend a considerable amount of effort in manually inspecting and refining the search results and looking for specific information or data of interest. The domain is challenging because of the inherent specificity of its language, rich in technical terms and syntactically complex structures, and because information on a specific topic is often issued by different PA bodies that are not connected with each other.
Innovation	This thesis will develop advanced tools for retrieving and structuring information extracted from documents issued by PA bodies. In particular, a "City Enabler" is an IT solution that offers a collaborative plaza allowing all relevant stakeholders (public and private bodies and well as citizens) to work together, both for publishing data and for exploiting them to create novel urban services. The proposed approach will extract and combine – in a flexible way – different information layers, ranging from the semantic content of documents to their similarity with other documents, coarse-grained topics, readability metrics, document "popularity", etc. This will lead to the development of a flexible information retrieval tool, in which search preferences are taken into account. In this way, different stakeholders will take advantage of the proposed solution, each having the possibility to look for specific types of information.

Research focus/topics	The proposed solution will combine research in information retrieval and in semantic processing of texts. The thesis will investigate different algorithms to rank search results and explore solutions to combine in such rankings the different information layers obtained from semantic analysis. The ranking algorithms will be evaluated by users and modified according to search preferences. From the semantic viewpoint, research will involve the analysis and comparison of different content extraction techniques in the PA context, comparing topic modeling, key-concept extraction, distributional approaches (e.g. word embeddings), etc. Research will mainly concern the best and most efficient integration between ranking algorithms and different layers of semantic analysis. The PA domain, with a well-defined yet extensive set of reference documents, is an ideal domain for research, since it is challenging from the linguistic and semantic point of view but offers a more controlled environment than standard general-purpose, search engine scenarios. Aside from this, it also offers the possibility of carrying out evaluations with "real" users and to directly measure the impact of different research solutions.
Expected outcome	The main expected outcome of the PhD thesis, from the application perspective, will be a software tool allowing for "intelligent" retrieval of information regarding the Public Administration domain, validated on a set of real use cases. Namely, such a tool should allow for the flexible, multi-layered searching of document content and rank the search output according to stakeholders' preferences. From the research perspective, the PhD work will result in a set of top-level publications in high-ranked conferences and journals, mainly in the areas of information retrieval and natural language processing.
Action Line	Digital Cities

Partnership

Industrial partner	Engineering Ingegneria Informatica S.p.A. - ENG (Italy)
Research partner	Fondazione Bruno Kessler - FBK (Italy)
HEI granting the title	ICT International Doctoral School – University of Trento
DTC location	Trento
PhD duration	3 years