

**Luis Miguel Sintra Salvo Paiva**

**Semantic relations extraction from unstructured information for domain ontologies enrichment**

Dissertação para obtenção do Grau de Mestre em

Engenharia Electrotécnica e de Computadores

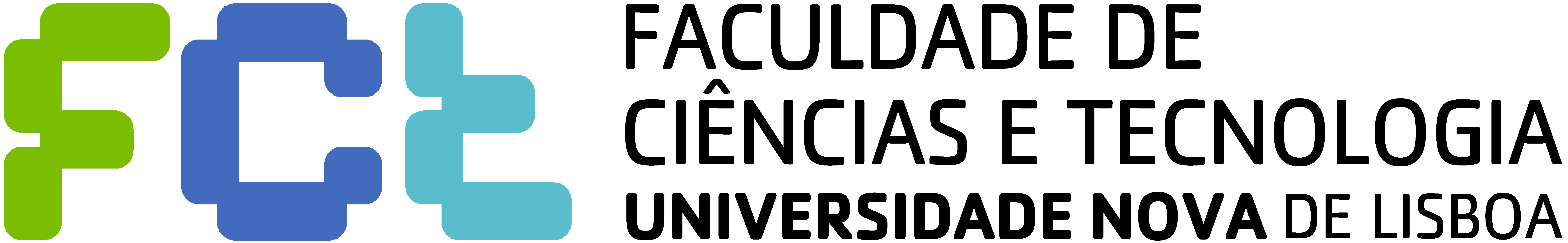
Orientador: Professor Pedro Maló, FCT-UNL

Co-orientador: Professor Celson Lima, MIT

Júri: *(Font: Arial, 10 pt normal)*

|  |  |
| --- | --- |
| Presidente: | Prof. Doutor(a) [Nome Completo] |
| Arguente(s): | Prof. Doutor(a) [Nome Completo] |
| Vogal(ais): | Prof. Doutor Pedro Maló |
|  | Prof. Doutor Celson Lima |
|  | Mestre Ruben Costa |

*(Tipo de letra: Arial, 10 pt normal)*



**Setembro 2014**



**Luis Miguel Sintra Salvo Paiva**

**Semantic relations extraction in unstructured information for domain ontologies enrichment**

Dissertação para obtenção do Grau de Mestre em

Engenharia Electrotécnica e de Computadores

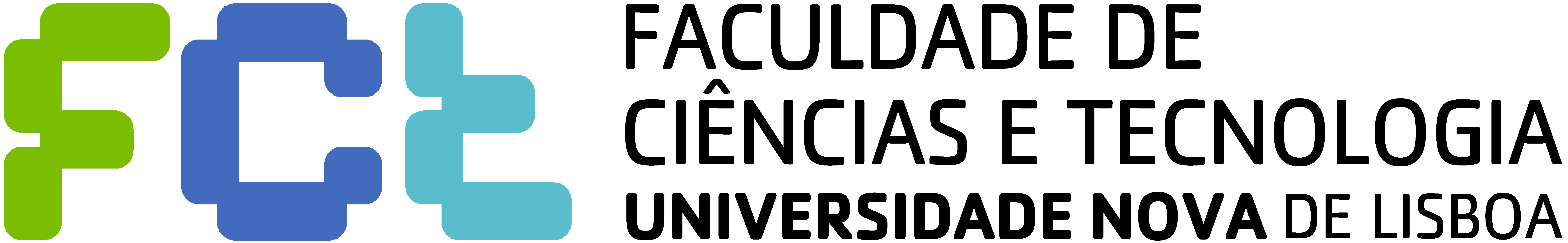
Orientador: Professor Pedro Maló, FCT-UNL

Co-orientador: Professor Celson Lima, MIT

Júri: *(Font: Arial, 10 pt normal)*

|  |  |
| --- | --- |
| Presidente: | Prof. Doutor(a) [Nome Completo] |
| Arguente(s): | Prof. Doutor(a) [Nome Completo] |
| Vogal(ais): | Prof. Doutor Pedro Maló |
|  | Prof. Doutor Celson Lima |
|  | Mestre Ruben Costa |

*(Tipo de letra: Arial, 10 pt normal)*



**Setembro 2014**

**COPYRIGHT**

**Semantic relations extraction in unstructured information for domain ontologies enrichment**

Luis Miguel Sintra Salvo Paiva – Todos os direitos reservados

A Faculdade de Ciências e Tecnologia e a Universidade Nova de Lisboa têm o direito, perpétuo e sem limites geográficos, de arquivar e publicar esta dissertação através de exemplares impressos reproduzidos em papel ou de forma digital, ou por qualquer outro meio conhecido ou que venha a ser inventado, e de a divulgar através de repositórios científicos e de admitir a sua cópia e distribuição com objectivos educacionais ou de investigação, não comerciais, desde que seja dado crédito ao autor e editor.

# Acknowledgements

As this was a long path to walk, I hope I can give the deserved merit to all that help me in some way to achieve this great and fulfilling goal.

# Abstract

(Ver em português para traduzir) Since semantic web appearance, several domain ontologies were developed and delivered in open access repositories. The existing domain ontologies describe semantic elements specific to a particular domain. These elements can be used complementing existing document information. This articulation can be duly empowered if, new methods could be created that in a semi automatic way, could help the ontologic precision. Specifically, the new patterns originating the building of new knowledge will be extracted not only from domain ontologies, but also from unstructured information sources.

Knowing the existence of scientific contributions in knowledge patterns extraction aimed to semantic enrichment, none considers the information inside an ontology. Desta forma, this document contributes with a method proposal of relations quantification between concepts from unstructured information, namely documents, and from knowledge existent in ontology to enrich it. To address this issue, this method is based on data mining techniques, specifically a technique to recognize semantic associations. This technique discovers data, namely concepts, that co-occurs from unstructured information.

To reach the proposed goal, a software support tool was developed, DOKES (*Dynamic Ontology Knowledge Enrichment Support*), that can be applied to any domain, however adapted to building and construction domain. DOKES main goal is to extract relevant knowledge from unstructured information and help its users enrich ontology.

Keywords: Artificial Intelligence, Semantic Web, Knowledge Discovery, Ontology Enrichment, Association Rules, Frequent Pattern

# Sumário

Desde o aparecimento da *web* semântica, várias ontologias de domínio foram desenvolvidas e disponibilizadas em repositórios de acesso aberto. As ontologias de domínio existentes pretendem descrevem elementos semânticos, como conceitos e as relações entre estes, específicos a uma determinada área. Elementos esses que podem ser usados como forma de complemento à informação existente em documentos. Esta complementaridade entre os elementos semânticos e os documentos, poderá ser devidamente potenciada, se forem criados novos métodos que, auxiliem o enriquecimento das ontologias. Mais especificamente, os novos padrões que dão origem a conhecimento novo ou enriquecido poderão ser extraídos não só de ontologias de domínio, como também de fontes de informação não estruturada.

Sabendo que existem bastantes contribuições ciêntíficas na área da extracção de padrões de conhecimento para enriquecimento semântico, destas nenhuma considera a informação contida na ontologia. Desta forma, este documento tem como principal objectivo contribuir com uma proposta de um método de quantificação de relações entre conceitos a partir de fontes de informação não estruturada, nomeadamente documentos, e do conhecimento já existente na ontologia, para enriquecimento das mesmas. De forma a desenvolver tal abordagem, este método tem como base, o recurso a técnicas de *data mining*, especificamente a uma técnica de reconhecimento de associações semânticas. Descrita como uma técnica para descoberta da co-ocorrência de dados, nomeadamente conceitos, a partir de informação não estruturada.

Para atingir o objectivo proposto, foi desenvolvida uma ferramenta de suporte a decisões para enriquecimento de conhecimento na ontologia, chamada DOKES (*Dynamic Ontology Knowledge Enrichment Support*), abrangente a qualquer domínio, mas adaptada ao sector da construção civil. O DOKES tem como principal objectivo a extracção de conhecimento relevante a partir de informação não estruturada e ajudar os seus utilizadores a enriquecer a ontologia.

Palavras-Chave: Inteligência Artificial, Web Semântica, Descoberta de Conhecimento, Enriquecimento de Ontologias, Regras de Associação, Padrões Frequentes

*Dedico a concretização desta etapa, finalizada por esta dissertação aos meus Pais, Mário Luiz e Maria Edite…*

*“Always look on the bright side of life!”*

*Monty Python, in “Life of Brian”*