Scuola universitaria professionale della Svizzera italiana
Dipartimento tecnologie innovative
Istituto sistemi informativi e networking

SUPSI

Using Apache Stanbol to automatically extract concepts from posts exchanged in Wordpress

The problem

Free work is a form of commercialization of activities that are directly functional to the productive system, receiving, in return, no (or very low) remuneration.

The project

The project aims to identify and distinguish the various forms of poorly paid or unpaid work and describe the concrete reality of the phenomenon in Ticino.

It leverages an online forum where users can discuss their experiences of free work.

An automatic system extracts concepts related to the free work domain from the discussed topics and store their statistics on a database.

The result

The online forum is implemented with Wordpress. The detection of forms of free work is done by sending exchanged posts to Apache Stanbol. The workflow comprises the following steps:

- A NLP phase (language detection, word tokenization);
- A Semantic Lifting process that links text processing to similar concepts (narrower, related, synonym) of our custom vocabulary;

An analytic dashboard shows statistics on the extracted concepts.

Contact information

Salvatore Vanini

Institute for Information Systems and Networking
salvatore.vanini@supsi.ch

Funding agency

Scuola Universitaria Statale della Svizzera Italiana (SUPSI)

Partners

Department of Business Economics, Health and Social Care
Department for Environment Constructions and Design
Institute for Information Systems and Networking

Research domain

Social systems and public health

Post in Wordpress

Sono riuscita ad ottenere un posto di **stage** da sogno … Che lo stage non fosse pagato all'inizio non mi aveva turbato … "Certo non la paghiamo, ma come **tirocinante** non starà qui a fare fotocopie o a preparare il caffè, siamo riconosciuti per essere degli ottimi formatori!" Le attività erano impegnative e trascorrevo in ufficio ben più di otto ore al giorno

Natural Language Processing

lang ID
Work classification Extraction

Apache Stanbol Enhancement Chain

Custom Dictionary

Keyword	Narrower	Synonym	Related
Stage	Stage di orientamento Stage di introduzione		Stagista Contratto di stage
Tirocinio			Tirocinante
Impiego remunerato			
Posto fisso			
Posto in organico			
Periodo di prova			
Programma/i di occupazione			Ufficio Regionale Collocamento URC Misure attive
Assicurazione e assistenza sociale	Attività di utilità pubblica Riqualifica professionale		Assicurazione Invalidità AI Non immediatamente occupabile

Semantic Lifting

Extracted concepts

Language	Concept	Related
Italiano	Stage	Tirocinante