

## Relazione esercitazione – FCA

L'obiettivo dell'ultima esercitazione è stato quello di applicare un metodo proveniente dall'ambito matematico chiamato **FCA (Formal Concept Analysis)** per fare Ontology Learning. L'Ontology Learning è un processo di apprendimento di un'ontologia in maniera automatica o semi-automatica. Esso può essere anche visto come quel processo che ci permette di poter passare da dati non strutturati (come quelli testuali) a dati strutturati in maniera automatica o semi-automatica. La FCA ci permette sostanzialmente di partire da dati non strutturati e ottenere automaticamente una tassonomia di concetti. Per poter fare questo la FCA parte da una matrice di adiacenza nella quale sono presenti come indici di colonna i nostri concetti (o comunque le entità del dominio considerato) e come indici di righe le features, che ci permettono di descrivere questi concetti. Ogni feature utilizzata è di tipo booleano e ciascuna di essa ci dice se un certo concetto ha o meno una certa proprietà. Dalla matrice di adiacenza, applicando la FCA, si ottiene quello che viene chiamato **contesto formale, che può essere rappresentato graficamente proprio come una tassonomia di concetti latenti**. Si parla di concetti latenti perché questi non erano espliciti nella matrice di adiacenza ma sono stati trovati e inseriti nella gerarchia tramite l'applicazione della FCA.

L'algoritmo implementato esegue i seguenti passi:

- 1) Innanzitutto, viene selezionata una frase casuale presente all'interno di un corpus che è stato creato appositamente per questa esercitazione. Esso è presente nella cartella chiamata "**Frase corpus**" ed è chiamato "**frase**". E' un piccolo corpus costituito da 20 frasi prese direttamente da alcune pagine di Wikipedia riguardanti argomenti come "animals", "football" e "Artificial Intelligence".
- 2) La frase ottenuta viene parserizzata mediante la chiamata della funzione "**get\_parser\_a\_dipendenze**" che, facendo uso della libreria **Spacy**, restituisce l'albero a dipendenze della frase iniziale. Dall'albero ottenuto vengono eliminate eventuali stopwords.
- 3) Dopodiché, vengono ottenute e memorizzate tutte le coppie [nome\_dipendenza, parola] che permettono di tener conto di quali sono i nomi delle dipendenze per ciascuna parola della frase.
- 4) A questo punto, utilizzando le coppie ottenute al passo precedente, viene costruita la matrice di adiacenza che avrà come indici di riga le parole di contenuto della frase considerata e come indici di colonna le features, che in questo caso saranno tutti i nomi delle dipendenze presenti nell'albero ottenuto al passo 2).
- 5) Infine, viene applicata la FCA, usando come input la matrice di adiacenza appena creata e utilizzando la libreria **Concepts** per ottenere il contesto formale.

## Risultati ottenuti

**Frase selezionata:** *“The expansion and cultural influence of the British Empire allowed these rules of football to spread to areas of British influence outside the directly controlled Empire.”*

**Matrice di adiacenza:**

	nsubj	amod	conj	compound	pobj	R00T	ccomp	prep	advmod
expansion	X								
cultural		X							
influence			X		X				
British		X		X					
Empire					X				
allowed						X			
rules	X								
football					X				
spread							X		
areas					X				
outside								X	
directly									X
controlled		X							

**Lattice (presente nella cartella “Risultati” con il nome di “lattice.pdf”:**

