

Linked Open Data per un Content-based Recommender System

Luciano Quercia
Simone Rutigliano

**Accesso intelligente alle informazioni ed
elaborazione del linguaggio naturale**

Corso di Laurea in Informatica Magistrale

30 aprile 2013



Outline

- 1 Obiettivi
- 2 Progetto
 - Sorgente dati
 - Realizzazione
 - Fattori
 - Output
- 3 Sperimentazione
 - Dataset
 - Protocollo Sperimentale
 - Risultati
- 4 Conclusioni e sviluppi futuri
 - Document Image Understanding



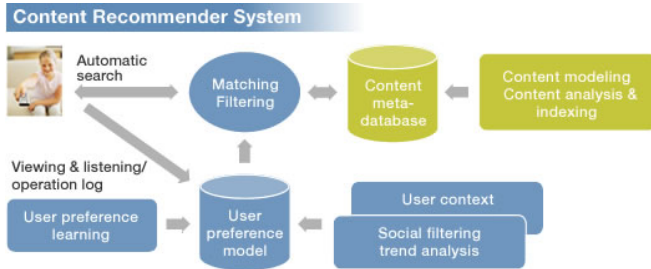
Obiettivi

- Realizzazione di un content-based recommender system
- Basato sulla Linked Open Data Cloud
- Utilizzando il formalismo RDF



Content-based Recommender System

Il sistema stabilisce a priori la distanza tra i film al fine di raccomandare i più simili alle preferenze dell'utente



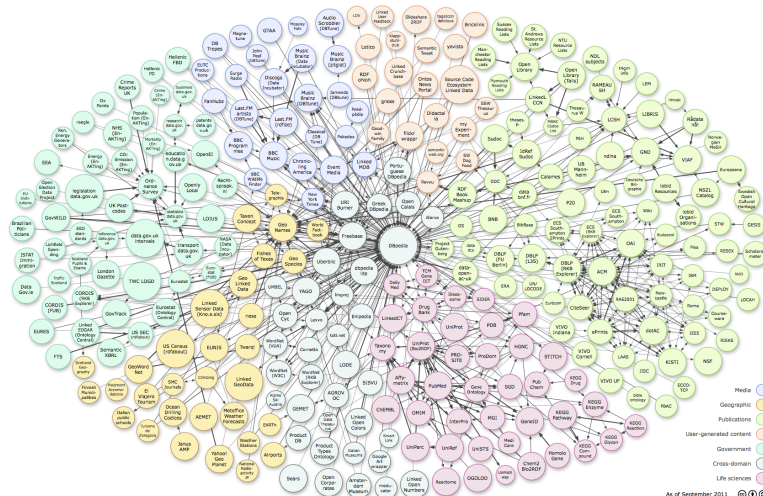
Linked Open Data

Il più grande *Cloud* al mondo di:

- dataset semantici
- fortemente interconnessi
- descritti attraverso RDF



Linked Open Data



As of September 2011



Resource Description Framework

Strumento base proposto da W3C per la codifica, lo scambio e il riutilizzo di metadati strutturati.

L'RDF Data Model si basa su tre principi chiave:

- 1 Qualunque cosa può essere identificata da un (URI)
- 2 utilizzare il linguaggio meno espressivo per definire qualunque cosa
- 3 Qualunque cosa può dire qualunque cosa su qualunque cosa



DBPedia

Fulcro della LODC Trasformazione in RDF di Wikipedia

Linked Data is about using the Web to connect related data that wasn't previously linked, or using the Web to lower the barriers to linking data currently linked using other methods. More specifically, Wikipedia defines Linked Data as a term used to describe a recommended best practice for exposing, sharing, and connecting pieces of data, information, and knowledge on the Semantic Web using URIs and RDF.



Proprietà estratte

Per la raccomandazione di film, abbiamo estratto le seguenti proprietà

- studio
- music
- music composer
- writer
- editing
- director
- subject
- starring
- productor
- writer
- cinematography



Grafo delle Risorse

Attraverso query SPARQL sono state estratte tutte le triple che avevano proprietà nota e un film come soggetto È stato generato il grafo delle risorse



Grafo dei Film

Tutte le risorse non film sono state epurate ed inglobate all'interno degli archi.



Distanze

Sono state applicate 4 distanze su grafo:

- Direct
- Combined
- Direct Weighted
- Combined Weighted



Rappresentazione del profilo

Il profilo è stato rappresentato in 2 modi:

Simple Insieme di film positivi per l'utente

Weighted Ogni film influisce, positivamente o negativamente, alle raccomandazioni, secondo il voto ricevuto



Raccomandazioni



MovieLens



Protocollo Sperimentale



Metriche



Risultati



Conclusioni e sviluppi futuri



TEPaC

TEPaC

Transductive Emerging Pattern based Classifier

- classificatore di strutture logiche
- basato su pattern emergenti
- utilizza un approccio trasduttivo



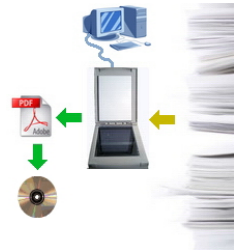
Document Image Understanding

- Comprensione automatizzata di documenti cartacei
- La maggior parte della conoscenza mondiale si trova su supporti cartacei
 - Libri
 - Documenti
 - Giornali
- La digitalizzazione offre innumerevoli vantaggi



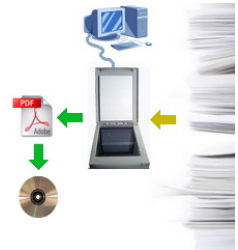
Document Image Understanding

- Comprensione automatizzata di documenti cartacei
- La maggior parte della conoscenza mondiale si trova su supporti cartacei
 - Libri
 - Documenti
 - Giornali
- La digitalizzazione offre innumerevoli vantaggi



Document Image Understanding

- Comprensione automatizzata di documenti cartacei
- La maggior parte della conoscenza mondiale si trova su supporti cartacei
 - Libri
 - Documenti
 - Giornali
- La digitalizzazione offre innumerevoli vantaggi



<i>minGR</i>	<i>minSup</i> (%)		
	30	40	50
1	528032	344798	254805
2	523274	341534	252355
8	516958	336733	248658
64	513503	334292	246843

Dataset TPAMI

<i>minGR</i>	<i>minSup</i> (%)		
	10	20	30
1	386996	176407	114492
2	382639	173372	112476
8	376645	169406	109814
64	374736	167742	108595

Dataset ICML

<i>minGR</i>	<i>minSup</i> (%)		
	10	20	30
1	128327	88684	58603
2	126840	87644	58091
8	122591	84208	55718
64	121363	82980	54490

Dataset BG



Grazie per l'attenzione.

