

# Linked Open Data for content-based recommender systems

Luciano QUERCIA  
Simone RUTIGLIANO

April 18, 2013

## 1 Introduzione

Il nostro progetto consiste nella realizzazione di un content-based recommender system che raccomanda film che utilizza i dati provenienti dalla Linked Open Data Cloud per hsdfbjsdhvfhfbsdjfhbsdufh

## 2 Lavori correlati

Mirizzi et al. (1), affrontando lo stesso problema, hanno utilizzato un VSM puro trasformando il grafo RDF di partenza in un tensore 3-dimensionale di adiacenza.

Passant (2), invece, in un contesto differente dal nostro (artisti musicali per poterle confrontare con le nostre

## 3 Progetto

## 4 Sperimentazione

## 5 Misure

### 5.1 Distanze

Il nostro recommender system è stato testato su quattro diverse misure di distanza. Le prime tre sono descritte in Passant (2), la quarta è stata realizzata *ex novo*.

### 5.1.1 Passant Direct

La distanza diretta. Indica il numero di archi diretti che collegano i film  $a$  e  $b$ . Tenendo conto che il grafo preso in considerazione è orientato, vanno presi gli archi in entrambe le direzioni.

In maniera formale,  $C_d(n, r_a, r_b)$  restituisce il numero di archi che vanno dal film  $a$  al film  $b$ .  $C_d(n, r_b, r_a)$  restituisce il numero di archi che vanno dal film  $b$  al film  $a$ .

La formula della distanza completa sarà quindi:

$$P_d(r_a, r_b) = \frac{1}{1 + C_d(n, r_a, r_b) + C_d(n, r_b, r_a)} \quad (1)$$

### 5.1.2 Passant Indirect

Indica la distanza indiretta tra i film presi in esame; il valore restituito sta ad indicare il numero di archi che mettono in relazione i due film passanti attraverso delle risorse condivise.

Formalmente, dati:

$$C_{io}(n, r_a, r_b) = |\{o \mid (r_a \text{ } p \text{ } o) \wedge (r_b \text{ } p \text{ } o)\}|, \text{ con } p \in P$$

$$C_{ii}(n, r_a, r_b) = |\{s \mid (s \text{ } p \text{ } r_a) \wedge (s \text{ } p \text{ } r_b)\}|, \text{ con } p \in P$$

la formula indiretta di Passant è:

$$P_i(r_a, r_b) = \frac{1}{1 + C_{io}(n, r_a, r_b) + C_{ii}(n, r_b, r_a)} \quad (2)$$

indica il numero di archi che collegano indirettamente una coppia di triple RDF (soggetto–predicato–oggetto), si sommano gli archi contenuti nelle triple RDF aventi come soggetti i due film in questione e come oggetto un film in comune ad entrambe le triple RDF.  $r_a$  – predicato –  $o_1$   $r_b$  – predicato –  $o_1$   $C_{ii}(n, r_a, r_b)$  indica il numero di archi che soddisfano la seguente proprietà: data una tripla RDF: soggetto – predicato – oggetto, si sommano gli archi contenuti nelle triple rdf aventi come oggetti i due film in questione e come soggetto un film in comune ad entrambi le triple RDF. oggetto1 – predicato –  $ra$  oggetto1 – predicato –  $rb$

### 5.1.3 Passant Combined

Indica la distanza combinata tra i due film in questione utilizzando sia la distanza indiretta, che quella diretta.

Nostra Si prendono in considerazione sia gli archi diretti che uniscono a e b, si gli archi indiretti, cioè quelli che legano a e b ad un oggetto o soggetto comune. Un link è un collegamento diretto tra due film con un suo peso dato dalla produttoria dei pesi degli archi precedentemente trovati. La distanza coincide con la sommatoria del peso di ogni link.

## 5.2 Profili

### 5.2.1 Simple

Il profilo coincide con l'insieme dei film valutati positivamente (4 e 5 considerati in egual misura)

### 5.2.2 Simple Negative

I film valutati 4 e 5 sono nell'insieme dei positivi. I film valutati 1 e 2 sono nell'insieme dei negativi.

### 5.2.3 Voted Nostra

Il profilo è l'insieme dei film visti, pesati secondo la formula: votazioni dei film visti - voto medio dei film.

### 5.2.4 Voted Musto

Il profilo è l'insieme dei film visti, pesati secondo la formula: Voto massimo + 1 - voto del film visto.

## References

- [1] R. Mirizzi, T. Di Noia, V. C. Ostuni, and A. Ragone. Linked open data for content-based recommender systems. 2012.
- [2] A. Passant. Measuring semantic distance on linking data and using it for resources recommendations. In *Proceedings of the AAAI Spring Symposium "Linked Data Meets Artificial Intelligence"*, volume 3, 2010.