

Organizzazione e contenuti della relazione di mini-progetto per il gruppo ACC

Alessandro Stefani¹, Caterina Buranelli², and Cristi Gutu³

¹ Corso di laurea in Statistica per le tecnologie e le scienze, matricola 1148387
`alessandro.stefani.6@studenti.unipd.it`

² Corso di laurea in Statistica per le tecnologie e le scienze, matricola 1234567
`caterina.buranelli@studenti.unipd.it`

³ Corso di laurea in Statistica per le tecnologie e le scienze, matricola 1147351
`gheorghecristi.gutu@studenti.unipd.it`

Sommario Questo progetto tratta la realizzazione attraverso il pacchetto *Whoosh*, di un motore di ricerca volto al reperimento di documenti della collezione sperimentale *OHSUMED* indicizzata opportunamente.

Keywords: *Information Retrieval · IR · statistica · reperimento · indicizzazione · Whoosh · Python*

1 Introduzione

L'obiettivo principale della relazione è da una parte la documentazione del progetto di un servizio di IR e dall'altra una misura del grado in cui si sia riusciti a mettere in pratica i contenuti della disciplina illustrati durante le lezioni.

A tal scopo la relazione dovrà illustrare nelle sezioni successive:

- i metodi di indicizzazione,
- i modelli di reperimento,
- l'interfaccia basata su un *browser* per il WWW
- i risultati della *valutazione* condotta con la collezione sperimentale OHSUMED.

Il lettore della relazione è lo studente medio di un corso di laurea in statistica al quale la relazione deve dare tutti gli strumenti per comprendere il contenuto. Ci si metta nei suoi panni e si scriva tutto ciò e solo ciò che serve. Chiedersi qual è il messaggio che lo studente deve “portarsi a casa”, esplicitarlo in questo paragrafo e concentrarsi su quello nel resto della relazione.

L'introduzione della relazione deve servire al lettore a capire se vale la pena continuare a leggere il resto. Si possono riassumere i contenuti delle sezioni successive e metterne in evidenza i punti principali. La relazione consiste di tre paragrafi principali dopo questa introduzione e prima della bibliografia, per la quale si suggerisce Bib_TEX se si scrive con L_AT_EX.

2 Base di partenza

La base di partenza è formata dai metodi documentati nei libri di testo. Si eviti di trascrivere pari pari, si cerchi piuttosto di rielaborare i contenuti in modo da renderli *coerenti* col resto della relazione; in particolare, si descrivano tutti e solo i metodi usati negli esperimenti e si eviti di parlare di quei metodi che poi non sono stati usati; ad esempio, se si conducono degli esperimenti con BM25F, si deve descrivere questo schema di pesatura in questa sezione.

3 Metodi proposti

Nel caso in cui si siano sviluppati:

- modelli di reperimento,
- metodi di indicizzazione,
- schemi di pesatura o
- altri metodi o tecniche

propri, non documentati in libri di testo o altra letteratura, si scriva in questa sezione una descrizione accurata e completa. Si mettano in evidenza le caratteristiche distintive dei propri contributi. Se non si è proposto nulla di nuovo, si scriva *Nessuno*. In una delle ultime lezioni si vedrà come implementare delle proprie funzioni di reperimento e schemi di pesatura.

4 Esperimenti

Si descriva la collezione sperimentale OHSUMED in termini di dimensione e tipo di dati. Si descrivano i risultati dei confronti tra metodi di base e/o quelli proposti. Cruciale è la descrizione accurata degli esperimenti: essa deve permetterne la replicazione.

Il *software*, sia quello a titolo esemplificativo dei concetti che quello che realizza i propri metodi, va caricato su moodle nella cartella messa a disposizione più avanti. Nella relazione va scritta la descrizione dell'algoritmo in modo preciso e completo da permetterne la riproduzione.

Riferimenti bibliografici

1. Autore, F.: Article title. Journal **2**(5), 99–110 (2016)
2. Autore, F., Autore, S.: Title of a proceedings paper. In: Editor, F., Editor, S. (eds.) CONFERENCE 2016, LNCS, vol. 9999, pp. 1–13. Springer, Heidelberg (2016). <https://doi.org/10.1007/1234567890>
3. Autore, F., Autore, S., Autore, T.: Book title. 2nd edn. Publisher, Location (1999)
4. Autore, A.-B.: Contribution title. In: 9th International Proceedings on Proceedings, pp. 1–2. Publisher, Location (2010)
5. LNCS Homepage, <http://www.springer.com/lncs>. Last accessed 4 Oct 2017



(a) Alessandro Stefanni



(b) Caterina Buranelli



(c) Cristi Gutu