

Presentazione del Modulo Acquisizione Automatica di Lessico e Terminologia

17 febbraio 2005

1 Informazioni generali

Docente Marco Baroni

Email baroni@sslmit.unibo.it

Telefono 0453/374744

Homepage del corso <http://www.e-learning.sslmit.unibo.it/TLS>

Ricevimento Lunedì 16:00-17:30 in aula Rosiello

2 Obiettivi Formativi

- Acquisizione di metodi per la rapida costruzione di risorse ausiliari alla traduzione/interpretazione settoriale.
- Capire meccanismi dietro a tali metodi, e dunque i loro punti deboli.
- Comunicare con informatici con cognizione di causa.
- Alfabetizzazione informatica.

2.1 Programma

- Il corso prosegue nel semestre autunnale.
- Visto il monte ore diverso, la quantità di materiale svolto in questo semestre vs. nel prossimo varierà tra interpreti e traduttori.
- Programma di massima:
 1. Introduzione generale al lavoro con i corpora.
 2. Cenni di probabilità e statistica.

3. Creazione di corpora dal web.
4. Tokenizzazione, part-of-speech tagging, lemmatizzazione.
5. Indicizzazione, consultazione e analisi di corpora.
6. Estrazione di collocazioni.
7. Estrazione di termini.

3 Materiali

- J. Peek, G. Todino and J. Strang. *Learning the UNIX Operating System, 4th Edition*. O'Reilly. 1998. Due copie (annotate da me) in portineria di via Oberdan.
- Handouts e articoli, che posterò in formato elettronico sul sito del corso.
- Metterò gli articoli non disponibili in formato elettronico in una teca nella portineria di via Oberdan.

3.1 Valutazione (10 punti in totale)

1. Test di abilità Unix: assegnato lunedì 21 febbraio, da consegnare entro lunedì 28 febbraio prima delle 17:30 (2 punti)
2. Compito a fine marzo: aspetti teorici, letture, lavoro svolto fino a quel punto (4 punti)
3. Compito a fine corso: aspetti teorici, letture, lavoro svolto (4 punti)

3.2 Da fare urgentemente

- Iscrivere a sito di e-learning.
- Fotocopiare *Learning the UNIX Operating System*.
- Formare gruppi di lavoro con argomento/lingua comune (lingue: italiano, inglese, tedesco, francese, spagnolo).
- Selezionare dominio specialistico su cui si vuole lavorare, e preparare file di testo con una decina/ventina di parole ritenute tipiche del dominio in questione.
- Ottenere una chiave per l'API di Google (istruzioni sul sito del corso).