



# Strumenti portabili per l'analisi di un forum in un ambiente di apprendimento online

**Stefano Penge**

stefano.penge@uniroma1.it

**Daniele Iori**

danieleiori77@gmail.it

Sapienza Università di Roma  
Facoltà di Scienze della Comunicazione



# Obiettivi

- Progettare, realizzare e sperimentare strumenti di analisi quantitativa di un forum all'interno di un ambiente di apprendimento che siano:
  - specifici per soggetti in apprendimento e non solo appartenenti ad una comunità virtuale
  - utilizzabili in tempo reale per l'autovalutazione di gruppo
  - disponibili per la comunità scientifica per facilitare un confronto tra contesti diversi



# Requisiti

- Gli strumenti realizzati dovranno quindi essere:
  - aperti
  - indipendenti dal contesto
  - rivolti non solo a ricercatori ma anche a tutor e a corsisti
- I dati prodotti dovranno essere
  - aperti
  - facilmente confrontabili



# Il quadro della ricerca

- Il campo di sperimentazione è dato dal monitoraggio del Corso Concorso per Dirigenti Scolastici 2006/2007 di PuntoEdu - ANSAS
- Direzione Scientifica: Prof. Mario Morcellini
- Team di ricerca: Ida Cortoni, Patrizia Cinti, Emmanuel Mazzucchi, M.Giovanna Onorati, Stefano Penge, Elena Valentini e altri



# Contesto

- Dimensioni
  - 2500 utenti
  - 21 moderatori
  - 24000 messaggi
  - 4000 topics
- Tecnologia
  - Invision PowerBoard
  - Oracle



## Strumenti offline?

- I testi di un forum non possono facilmente essere analizzati con strumenti offline di analisi lessicale:
  - Non costituiscono un corpus unico ma una raccolta di microtesti
  - Hanno una forte dipendenza dal tempo e dalla relazione (sequenza, topic)
  - Per i tutor è utile un'analisi in tempo reale dei dati per apportare correttivi e per bilanciare l'uso di altri strumenti di comunicazione/collaborazione



# Strumenti online

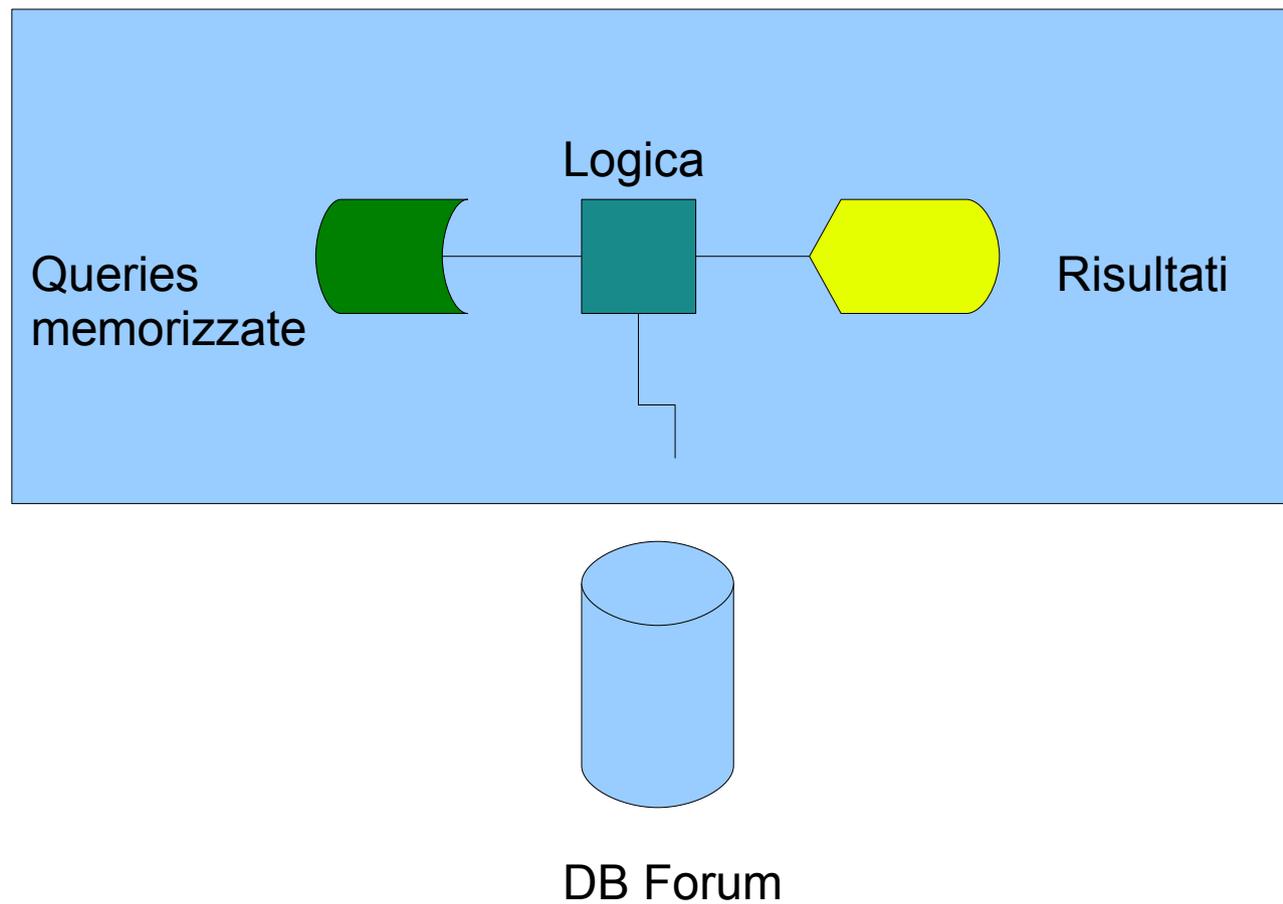
- Di qui l'idea di sviluppare un insieme di strumenti online applicabili direttamente al database del forum:
  - accessibili via web
  - di uso semplice
  - modulari / estensibili



# Tools

- Query SQL standard, parametrizzate
- Moduli PHP per esecuzione e visualizzazione delle query
- Fogli di calcolo per elaborazione dei risultati, grafici
- Indicazioni per l'uso (normalizzazione, ...)

# Tools





# Modelli di analisi

- Esistono ormai, in Italia e all'estero, diversi modelli di analisi quantitativa e qualitativa dei post di un forum
  - Simoff, Fafchamps, Levin, Jeong...
- Anche in Italia ci si occupa in tanti del tema
- Sarebbe interessante poter confrontare i risultati...



# Tipi di analisi

- Per le dimensioni del contesto, un'analisi qualitativa non sarebbe possibile, o non soddisferebbe alcuni requisiti (semplicità d'uso, tempo reale)
- Le analisi quantitative hanno senso in un quadro teorico più generale, che si può costruire solo se i dati sono confrontabili con altri contesti



# Confronti

- Ad esempio, nel caso del monitoraggio del CC a noi interessa valutare l'intero ambiente, più che i corsisti, in modo da poterlo confrontare con altri ambienti simili :
  - altre edizioni del Corso Concorso
  - altri ambienti dentro PuntoEdu
  - altri forum che usano Invision Power Board
  - altri forum...



# Varabili

- A partire da modelli esistenti, occorre selezionare delle variabili che siano definibili chiaramente e indipendentemente dalle condizioni specifiche, e che siano misurabili con strumenti trasversali



# Variabili 1

- Il Modello Open University prende in esame cinque categorie di variabili:
  - a) numero totale dei messaggi (indice della partecipazione generale);
  - b) numero dei messaggi degli studenti in relazione al numero di messaggi dei tutor (indice della partecipazione attiva);
  - c) numero di messaggi per studente (indice delle attività degli studenti nella classe virtuale);
  - d) numero di messaggi in un dato arco di tempo (indice dell'andamento diacronico della partecipazione);
  - e) lunghezza dei messaggi in un dato arco di tempo (indice dell'andamento del livello di interesse).



## Variabili 2

Sarebbe utile definire con precisione gli elementi sulla base dei quali istituire un confronto tra forum. Per esempio::

- **numero** di post: totale, medio per corsista, medio per giorno
- **dimensioni** dei post: totale, medio per corsista, medio per giorno.

Dove il tipo di forum non consente l'individuazione di threads e subthreads; è possibile però calcolare

- la *vivacità* media dei thread (numero di post nell'unità di tempo)
- la *partecipazione* (numero di post per thread)
- la presenza di *leader* (numero di post dello stesso autore)



# Variabili 3

Un altro aspetto interessante potrebbe essere il rapporto tra il lavoro dei moderatori e quello dei corsisti:

- rapporto moderatori / corsisti
- rapporto post moderatore / post corsista
- frequenza di post giornalieri dei moderatori
- presenza dei post dei moderatori nei thread più vivaci
- creazione da parte dei moderatori dei thread più significativi



# Il caso Corso Concorso

- Nel nostro caso, non è stato possibile accedere direttamente al DB di Invision
- I dati sono stati esportati in formato dump di Oracle, poi importati e normalizzati all'interno di un DMBS OpenSource (MySQL).
- Qui vengono analizzati con delle query SQL (circa 60)



# Fasi di lavoro

In dettaglio si è trattato di:

- effettuare l'ordinamento dei temi (termini chiave usati nei subject dei messaggi) per frequenza, distribuiti per giorno
- effettuare l'ordinamento dei termini usati nel testo per frequenza, distribuiti per giorno
- analizzare la distribuzione di alcuni termini specifici (indicatori)
- confrontare le frequenze degli indicatori nelle classi di corsisti sopra individuate
- analizzare le frequenze degli indicatori in rapporto alla media dei post giornalieri e alla media delle dimensioni dei post



# Risultati

- E' in corso di stesura la relazione finale con i dati relativi al lessico usato nel Corso Concorso per Dirigenti Scolastici
- Insieme alla relazione verranno resi pubblici gli strumenti usati e le indicazioni per l'uso
- I moduli e le query verranno rilasciate con licenza GPL



# Queries: 1

- **Analisi attività moderatori**
- a. Lista dei moderatori ordinata per quantità di interventi e con il numero di forum in cui sono moderatori
- b. Lista dei moderatori ordinata per media di intervento nei forum ricavata dal numero totale di azioni rispetto al numero di forum in cui sono moderatori
- c. Lista dei moderatori ordinata per quantità dei post con indicazione del numero dei forum che modera
- d. Lista dei moderatori ordinata per media dei post nei forum ricavata dal numero totale di post rispetto al numero di forum in cui sono moderatori



# Queries: 2

- Analisi quantitativa degli utenti divisa per: utenti totali | utenti tranne moderatori | moderatori
  - a.Numero utenti
  - b.Numero utenti divisi per sesso
  - c.Numero utenti divisi per regione
  - d.Numero utenti divisi per Regione e sesso
  - e.Numero utenti divisi per Provincia
  - f.Numero utenti divisi per Provincia e sesso
  - g.Tutte le statistiche elencate fino ad ora ripetute per fascia d'età



## Queries 3 - 9

- Medie sull'età degli utenti
- Analisi quantitativa sui post
- Distribuzioni di frequenza dei post
- Analisi sulla lunghezza dei post
- Medie topic, rapporto post/topic
- Analisi quantitativa dei topic
- Medie sui topic



# Esempi

	Utenti	Corsisti	Moderatori
MEDIA	134,16	127,56	6,91
MODA	109	141	4
F	85,4	84,07	2,61
M	47,91	44,02	5,05

	Utenti	Corsisti	Moderatori	M/C
Media battute	1084,6912	1115,9286	773,8925	0,69
F	1110,1224	1130,8178	685,2168	0,61
M	1047,3526	1093,7553	795,8438	0,73



## Modelli: 2

- In altri contributi abbiamo proposto un modello di analisi lessicale basato sul concetto di evoluzione del linguaggio di un gruppo in apprendimento come indicatore dell'andamento del gruppo stesso
- Abbiamo cominciato ad applicare questo modello per la prima volta sul forum CCDS



# Evoluzione

<b>FASE A (iniziale)</b>		<b>FASE Z (finale)</b>
l'interazione linguistica utilizza termini impropri, presi dal linguaggio comune	REGISTRO	il discorso utilizza termini specialistici corretti
ha come oggetto le difficoltà comunicative	OGGETTO	riflette sul lavoro fatto insieme
si concentra sul linguaggio, sui termini anziché sui fatti	LIVELLO	si concentra sulle regole e sulla loro applicazione
procede per richieste continue	MODALITA'	propone nuovi elementi informativi
esprime dubbi su successo e sul profitto del processo formativo	TONALITA'	esprime una soddisfazione generale per le attività svolte



## Q 10: analisi lessicali

- Analisi delle occorrenze delle parole: Scuola, tecnologie, Docenti, Nuove Tecnologie, Insegnanti, apprendimento, Tic, Computer, Dirigente, Sicurezza, Informatica, Didattica; e delle combinazioni: Computer-Apprendimento, Computer-Scuola, Dirigente-Computer, Dirigente-Scuola, Dirigente-Tecnologie, Docenti-Computer, Informatica-Didattica, Tecnologie-Apprendimento, Tecnologie-Didattica, Tecnologie-Scuola.
  - a. Occorrenze divise per provincia e sesso
  - b. Tutte le statistiche elencate fino ad ora ripetute per fascia d'età



## E adesso?

- La proposta è quella di creare un gruppo di lavoro trasversale, per generalizzare gli strumenti sulla base dei dati di altri forum in altri contesti di apprendimento
  - sperimentare i tools su diversi forum,
  - testare gli indicatori,
  - confrontare i risultati



# Per collaborare:

Stefano Penge

[stefano.penge@uniroma1.it](mailto:stefano.penge@uniroma1.it)