



Per il ciclo di incontri "Nexa Lunch Seminar"

## 7° Nexa Lunch Seminar

### L'intelligenza collettiva per organizzare la conoscenza sul Web

<http://nexa.polito.it/lunch-7>

**Mercoledì 23 gennaio 2012, ore 13 – 14**  
(4° mercoledì del mese)

**Nexa Center for Internet & Society (DAUIN)**  
**Via Boggio 65/a, Torino (primo piano)**

**Ingresso libero**

**Live streaming:** <http://nexa.polito.it/upcoming-events>

*Saranno disponibili panini e bibite per coloro che si saranno registrati su*  
<http://lunch07.eventbrite.com/> entro il 21 gennaio

L'organizzazione della conoscenza presente sul Web è uno degli strumenti principali per ridurre l'*information overload* sperimentato quotidianamente dagli utenti di Internet. In occasione del 7° Nexa Lunch Seminar **Federico Cairo**, research fellow del Centro Nexa su Internet & Società, illustrerà il suo lavoro di ricerca, riguardante la **classificazione automatizzata dei documenti in Rete per mezzo dei concetti presenti su DBpedia** (<http://dbpedia.org/>). Nato nel 2007 dalla collaborazione di Università Libera di Berlino (<http://www.fu-berlin.de/en/>), Università di Lipsia (<http://www.zv.uni-leipzig.de/>) e OpenLink Software (<http://www.openlinksw.com/>), DBpedia è un progetto per l'estrazione di dati strutturati dal "corpus" di Wikipedia e alla pubblicazione dei dati stessi sul Web in formato Linked Data (<http://linkeddata.org/>).



L'opportunità dell'utilizzo di DBpedia per la classificazione della conoscenza sul Web è sostenuta sia da aspetti puramente tecnici sia da valutazioni più teoriche. Essendo DBpedia collegata a un vasto "corpus" multilingue preannotato di carattere enciclopedico (Wikipedia), essa risulta tecnicamente molto adatta ad essere utilizzata per procedimenti automatici di natural language processing ([en.wikipedia.org/wiki/Natural\\_language\\_processing](http://en.wikipedia.org/wiki/Natural_language_processing)) e di text mining ([en.wikipedia.org/wiki/Text\\_mining](http://en.wikipedia.org/wiki/Text_mining)). Ma la cosa più interessante è che i concetti presenti in DBpedia sono il risultato di un **consenso semantico raggiunto in maniera collaborativa** dalla comunità degli internauti. Un criterio efficace di classificazione sul Web, infatti, non può essere imposto dall'alto, ma deve seguire gli stessi principi di libertà e trasparenza che hanno da sempre costituito l'essenza di Internet.

La conoscenza contenuta in Wikipedia, e di conseguenza in DBpedia, è frutto di quell'intelligenza collettiva e di quella cultura collaborativa che sembrano emergere come i tratti costitutivi delle comunità in Rete. I software che utilizzano Wikipedia e DBpedia per l'estrazione e l'arricchimento dei contenuti sul Web, come per esempio TellMeFirst (<http://tellmefirst.polito.it/>), progetto vincitore di un grant Working Capital Telecom 2011 (<http://www.workingcapital.telecomitalia.it/2011/10/tellmefirst/>) e ulteriormente sviluppato all'interno del Centro Nexa da Federico Cairo e Giuseppe Futia, fanno appunto leva su questa "saggezza della folla", mostrandone le potenzialità e i benefici per gli utenti di Internet.

<http://nexa.polito.it/contact>

## Biografia:

**Federico Cairo** (<http://nexa.polito.it/people/fcairo>), research fellow presso il Centro Nexa su Internet & Società e dottorando di ricerca al Dipartimento Interuniversitario di Studi regionali (DIST) del Politecnico di Torino, si è laureato in Filosofia presso il Collegio Superiore dell'Università di Bologna (2005) e in Ingegneria dell'Informazione presso il Politecnico di Torino (2008). La sua attività di ricerca si concentra principalmente sulle tecnologie del Semantic Web per gli archivi digitali e su motori di ricerca basati su ontologie. Dopo aver collaborato come ontology engineer con CSI-Piemonte e CELI sui progetti CoOPERARE (2010), dati.piemonte.it (2010), OPSA (2011), è stato uno dei vincitori dell'edizione 2011 di "Working Capital - Premio Nazionale Innovazione" di Telecom Italia con il progetto TellMeFirst, un'applicazione di arricchimento dei testi via Linked Open Data.

## Letture consigliate:

- Exner, P., Nugues, P. (2012) *Entity extraction: From unstructured text to DBpedia RDF triples*. The Web of Linked Entities Workshop (WoLE 2012)
- Mendes, P. N., Jakob, M., García-Silva, A., & Bizer, C. (2011). *DBpedia Spotlight: Shedding Light on the Web of Documents*. Text (pp. 1-8). Facultad de Informática (UPM)
- Muñoz-García, O., García-Silva, A., Corcho, Ó., De La Higuera Hernández, M., & Navarro, C. (2011). *Identifying Topics in Social Media Posts using DBpedia*. (J.-D. Meunier, H. Hrasnica, & F. Genoux, Eds.)Media, 28020, 81-86. Facultad de Informática (UPM)
- Bizer, C., Lehmann, J., Kobilarov, G., Auer, S., Becker, C., Cyganiak, R., & Hellmann, S. (2009). *DBpedia - A crystallization point for the Web of Data*. Web Semantics Science Services and Agents on the World Wide Web, 7(3), 154-165. Elsevier.

## Che cosa sono il Centro Nexa su Internet & Società e i Nexa Lunch Seminar

Il Centro Nexa su Internet & Società (<http://nexa.polito.it/>) del Politecnico di Torino (Dipartimento di Automatica e Informatica) nasce a partire dalle attività di un gruppo di lavoro multidisciplinare – tecnico, giuridico ed economico – formatosi a Torino nel 2003 e che da allora ha ideato, progettato e realizzato diverse iniziative, tra cui: Creative Commons Italia (2003-presente), CyberLaw Torino (2004), Harvard Internet Law Program Torino (2005), SeLiLi, servizio di consulenza legale gratuita sulle licenze aperte per creativi e programmatori (2006-presente), COMMUNIA, la rete tematica europea costituita da 50 partner sul pubblico dominio digitale (2007-2011) e LAPSI, la rete tematica europea dedicata agli aspetti legali delle informazioni del settore pubblico (2010-2012).

I "Nexa Lunch Seminar" (<http://nexa.polito.it/lunch-seminars>) costituiscono un ciclo di incontri proposto dal Centro Nexa su Internet & Società e si svolgono ogni quarto mercoledì del mese dalle ore 13 alle ore 14. Ogni secondo mercoledì del mese continua il consueto appuntamento con i "Mercoledì di Nexa" (<http://nexa.polito.it/mercoledi>): il prossimo incontro si svolgerà **mercoledì 13 febbraio 2013** alle ore 18 e avrà come ospiti **Federico Morando**, managing director del Centro Nexa su Internet & Società e lead di Creative Commons Italia, e **Massimo Travostino**, avvocato, fellow del Centro Nexa e membro del gruppo di lavoro giuridico di Creative Commons Italia.

## Responsabile Comunicazione Centro Nexa su Internet & Società:

### Giuseppe Futia

Tel: +39 011 090 7219

Mob: +39 333 1103017

[giuseppe.futia@polito.it](mailto:giuseppe.futia@polito.it)

<http://nexa.polito.it>

Twitter: <http://twitter.com/nexacenter>

Facebook: <http://facebook.com/nexa.center>

<http://nexa.polito.it/contact>