

## SEMINARI TECNICI FORMATIVI - marzo-dicembre 2014

Migrazione di tecnologie e linguaggi.

Teorie e pratiche per il trattamento di testi nelle scienze umane

[www.iliesi.cnr.it/seminari\\_formativi](http://www.iliesi.cnr.it/seminari_formativi)

**13 marzo 2014 - Ada Russo, ILIESI-CNR**

### **Teorie e pratiche nell'uso dell' XML/TEI per la codifica dei testi**

L'ILIESI ha adottato lo standard di codifica XML-TEI nel 2006, quando l'Istituto ha partecipato come content-partner al progetto europeo "Discovery. Digital Semantic Corpora for Virtual Research in Philosophy". L'adozione relativamente recente di TEI si accompagna però ad una lunga pratica dell'ILIESI nel campo della codifica dei testi. L'attenzione al testo, ai singoli termini e alle famiglie lessicali rappresenta il nucleo intorno a cui si sviluppano le ricerche che l'Istituto conduce sulla storia delle idee e della terminologia di cultura. Il seminario si articolerà in tre parti: la prima sarà dedicata ad una breve introduzione su codifica e linguaggi di markup; successivamente verrà illustrato il linguaggio XML-TEI, con particolare riguardo al suo impiego per la codifica di testi a stampa in edizioni d'epoca, in vista di elaborazioni lessicografiche e per la costituzione di corpora testuali; infine verranno descritti gli strumenti e le procedure per l'applicazione della codifica e la pubblicazione dei testi.

---

**9 aprile 2014 - Roberto Vitali, Sapienza, Università di Roma**

### **Latex: l'arte di formattare testi**

LaTeX è uno strumento per la composizione tipografica di alta qualità molto diffuso tra la comunità scientifica e accademica, soprattutto nelle branche delle scienze matematiche, fisiche e naturali. Il maggior punto di forza di LaTeX risiede nel disaccoppiamento tra il contenuto di un testo e la sua presentazione. Questo permette all'autore di concentrarsi sul contenuto durante la stesura di un documento e ottenere allo stesso tempo un risultato finale di ottimo livello.

Il seminario vuole illustrare le potenzialità di LaTeX e far emergere come il suo utilizzo possa arricchire, sia per la qualità del prodotto finale, che per la maggior produttività, anche la composizione di testi legati alle scienze umanistiche. Il seminario sarà strutturato in una prima parte di introduzione a LaTeX, successivamente saranno presentati i dettagli della composizione tipografica e infine verrà mostrato come realizzare un documento tramite LaTeX.

---

## SEMINARI TECNICI FORMATIVI - marzo-dicembre 2014

Migrazione di tecnologie e linguaggi.

Teorie e pratiche per il trattamento di testi nelle scienze umane

[www.iliesi.cnr.it/seminari\\_formativi](http://www.iliesi.cnr.it/seminari_formativi)

### **19 maggio 2014 - Fabrizio Pecoraro, IRPPS**

#### **Piattaforme per pubblicazioni online**

Lo sviluppo di modelli economici sostenibili basati su tecnologie open source per la gestione e la diffusione delle attività editoriali rappresenta una grande opportunità per gli Istituti di ricerca per migliorare la diffusione delle loro informazioni scientifiche. L'ampia espansione dell'editoria elettronica sta portando le biblioteche a svolgere un ruolo chiave nel miglioramento di sistemi innovativi per la divulgazione dei risultati della ricerca scientifica, con l'obiettivo di aumentare la qualità dei prodotti e ridurre i costi delle pubblicazioni. Questo aspetto è particolarmente evidente considerando la grande diffusione di riviste ad accesso aperto pubblicate da istituzioni accademiche ed enti di ricerca.

Obiettivo del seminario è illustrare le potenzialità degli strumenti forniti dalle tecnologie dell'informazione con particolare riferimento ai *Journal Management System* nella gestione in-house delle riviste scientifiche e nella riorganizzazione delle attività editoriali. L'attenzione verrà focalizzata sul software open source Open Journal Systems (OJS) di cui verranno descritte le principali funzionalità, sia dal punto di vista tecnico che editoriale, evidenziando vantaggi e problematiche nel suo utilizzo nell'intero processo editoriale di una o più riviste scientifiche. Verrà infine presentato un caso di studio sulla base dell'esperienza maturata all'interno dell'Istituto di Ricerche sulla Popolazione e le Politiche Sociali del Consiglio Nazionale delle Ricerche (IRPPS-CNR) nello sviluppo del servizio *CNR-IRPPS e-Publishing* per la pubblicazione dei prodotti editoriali dell'Istituto.

---

### **19 giugno 2014 - Alessio Piccioli, Net7**

#### **Muruca: strumenti e applicazioni per la gestione di biblioteche digitali semantiche**

Muruca è un framework, costituito da un insieme di strumenti open source, sviluppato per la realizzazione di biblioteche digitali semantiche.

Partecipando a numerosi progetti nazionali ed europei Net7, a partire dal 2001, ha realizzato Muruca che consente di gestire e di rendere fruibili oggetti digitali, di archiviare fonti primarie e secondarie (manoscritti, lettere, immagini di dipinti, spartiti

## **SEMINARI TECNICI FORMATIVI - marzo-dicembre 2014**

Migrazione di tecnologie e linguaggi.

Teorie e pratiche per il trattamento di testi nelle scienze umane

[www.iliesi.cnr.it/seminari\\_formativi](http://www.iliesi.cnr.it/seminari_formativi)

musicali, fotografie, video) in molteplici formati (xml, doc, odt, pdf, tiff, gif, jpeg, png, mpeg, avi, flv).

Muruca permette di navigare tra i contenuti digitali attraverso l'utilizzo di tag semantici, include, inoltre, strumenti progettati per arricchire semanticamente i contenuti che sono nativamente interoperabili con progetti come Europeana, Iconclass e Bibserver.

Il seminario sarà dedicato all'uso, alla modulazione e alla gestione di Muruca con particolare attenzione all'analisi dei principi alla base della costruzione di una biblioteca digitale e al significato del termine "semantico".

---

**16 ottobre 2014 - Felice Dell'Orletta, ILC**

**La piattaforma T2K: dal testo alla conoscenza**

La necessità quotidiana di accedere a grandi quantità di dati per lo più testuali ha dato grande impulso allo sviluppo di tecnologie per l'acquisizione, la classificazione e la gestione automatica del contenuto testuale e al loro sempre più diffuso impiego in una miriade di contesti applicativi.

T2K è una piattaforma web finalizzata all'acquisizione di informazione semantico-lessicale da corpora di dominio. Attraverso l'uso combinato di tecniche statistiche e di strumenti avanzati di Trattamento Automatico del Linguaggio, T2K è in grado di analizzare il contenuto dei documenti, estrarre i termini e le entità (nomi, luoghi, persone) potenzialmente più significativi, individuare le relazioni che legano termini e entità, indicizzare il corpus di partenza e ricostruire una mappa multidimensionale della conoscenza di dominio contenuta nella collezione documentale.

---

**13 novembre 2014 - Marco Veneziani, ILIESI**

**La lessicografia dei testi di cultura. Un esempio di software dedicato**

Tra la fine dell'Ottocento e la prima metà del Novecento hanno visto la luce molte grandi edizioni di classici della storia delle idee: Descartes, Spinoza, Hobbes e Locke, Bruno,

## SEMINARI TECNICI FORMATIVI - marzo-dicembre 2014

Migrazione di tecnologie e linguaggi.

Teorie e pratiche per il trattamento di testi nelle scienze umane

[www.iliesi.cnr.it/seminari\\_formativi](http://www.iliesi.cnr.it/seminari_formativi)

Galileo e Vico. Frutto di una filologia per alcuni aspetti invecchiata, sono ancora oggi quelle maggiormente citate negli studi. Per altre opere e per altri autori si è preferito ripubblicare in edizioni anastatiche i migliori testi d'epoca, mentre numerosi e meritevoli scritti attendono ancora di essere rimessi in circolazione.

Si tratta di un corpus variegato, redatto in latino e nelle diverse lingue europee, che per il suo valore storico, nonché per il segno che ha lasciato nella cultura di oggi, merita una lettura critica il più possibile aderente ai contenuti concettuali, e – se necessaria – una resa adeguata nelle lingue moderne.

Nello studio analitico di queste opere, grande può essere l'apporto delle recenti tecnologie informatiche, dai semplici programmi di scrittura alle applicazioni web più complesse e sofisticate, fino alla possibilità di un esteso ricorso agli spogli lessicografici e lessicometrici. Il seminario vuole sottolineare le nuove opportunità tecniche e presentare alcune effettive realizzazioni.

---

**11 dicembre 2014 - Aldo Gangemi, ISTC**

**Il *machine reading* come metodo per collegare gli archivi testuali al Web Semantico**

Il machine reading è un paradigma di estrazione dell'informazione (*Open Information Extraction*) che si applica in generale (cioè non è vincolato a un dominio) e non ha bisogno di esempi su cui addestrare un classificatore (cioè non è supervisionato).

Recentemente, la ricerca sul Web Semantico e l'ingegneria della conoscenza ha cominciato a ibridarsi con gli strumenti e le risorse sviluppate dalla comunità di trattamento delle lingue naturali, sviluppando nuovi strumenti che aggiungono al paradigma del machine reading anche la capacità di estrarre l'informazione dal testo, formalizzandola ed eseguendo inferenze automatiche. Si parla dunque di *estrazione della conoscenza (Open Knowledge Extraction)*.

Il seminario descriverà questa evoluzione della ricerca e come gli strumenti attuali possono essere usati per la produzione di dati strutturati e collegati a dati esistenti (*Linked data*) a partire da archivi testuali. Il problema di ragionare automaticamente su una grande quantità di conoscenza estratta automaticamente da testi fra loro collegati in serie e/o dinamicamente sarà anche oggetto di discussione.