



Deliverable D2.3

Sistemi di Annotazione Semantica di Contlets

Responsabile:	Di Martino Beniamino
Afferenza	Seconda Università di Napoli
Autori	Prof. Rocco Aversa, Ing. Carlo Baia, Prof. Ida Caracciolo, Prof. Beniamino DI Martino, Prof. Pasquale Femia, Ing. Angelo Martone, Ing. Francesco Moscato, Prof. Francesco Palmieri, Ing. Massimiliano Rak, Ing. Gianmarco Romano, Ing. Salvatore Venticinque, Prof. Rosanna Verde
Afferenza	Seconda Università di Napoli
Autori	Prof. Paola Velardi
Afferenza	Università di Roma "Sapienza"
Autori	Paolo Alongi, Raffaella Lantini
Afferenza	SPACE
Autori	Dr.ssa Federica Pesce, Dr. Fabrizio Melorio
Afferenza:	Istituto Luigi Sturzo

PROGETTO LC3	Revisione n*	0	Del	----
--------------	--------------	---	-----	------



INDICE

- [TR2.3.1 Stato dell'arte dei modelli ed Analisi di Sistemi di annotazione basati su ontologie](#)
- [TR2.3.2 Installazione, analisi e confronto funzionale dei sistemi di annotazione basati su ontologie](#)
- [TR2.3.3 Realizzazione di un'architettura per l' annotazione semantica e di uno strumento Prototipale](#)
- [TR2.3.4 Stato dell' arte ed Analisi delle metodologie e delle tecniche per la Classificazione ed il Clustering](#)
- [TR2.3.5 Stato dell' arte ed Analisi delle tecnologie e degli strumenti per la Classificazione Documentale](#)
- [TR2.3.6 Realizzazione di uno strumento prototipale per l' annotazione semantica automatica di corpora documentali basato su tecniche di classificazione automatica](#)
- [TR2.3.7 Architettura generale e intercambio dei dati semantici tra il MAG Repository e lo strumento di annotazione semantica](#)
- [TR2.3.8 Supporto alla definizione delle specifiche funzionali del Sistema MAG Teca](#)

Lo scopo di questo task è stato quello di analizzare e confrontare i modelli ed i sistemi disponibili per effettuare l'annotazione semantica (manuale o parzialmente automatizzata) di documenti multimediali e per effettuare la classificazione automatica degli stessi.

Nel seguito si descrivono le attività svolte ed i risultati conseguiti dalle Unità di ricerca della Seconda Università di Napoli (DII e JM).

E' stato analizzato lo stato dell'arte dei modelli e dei sistemi di annotazione semantica, con particolare riferimento all' analisi delle metodologie di annotazione e dei relativi criteri di classificazione, dei tipi di risorsa annotata, dei destinatari dell' annotazione, dei livelli di formalità e dei livelli di restrizione terminologica imposti sul linguaggio, del posizionamento dell' annotazione.

E' stata effettuata un'attività di analisi e confronto delle metodologie e tecniche per l'annotazione automatica, manuale ed automatizzata, di *contlets*. Una rassegna di strumenti per l'annotazione semantica è stata effettuata al fine da verificare quali tecniche e metodologie sono presenti in letterature ed è in fase di attuazione un confronto fra tutte

PROGETTO LC3	Revisione n*	0	Del	-----
--------------	--------------	---	-----	-------



quelle individuate al fine di individuare una nuova tecnica per l'annotazione adatta ai fini definiti dal progetto.

Sono stati analizzati e confrontati i principali strumenti Open Source per l'annotazione basati su Ontologie, tra i quali OntoMat-Annotizer, MnM, Smore, Cohse, Trellis, Melita, Gate, KIM, Interop KMap. Vedi Technical Reports **TR2.3.1 Stato dell'arte dei modelli ed Analisi di Sistemi di annotazione basati su ontologie** e **TR2.3.2 Installazione, analisi e confronto funzionale dei sistemi di annotazione basati su ontologie**

Si è progettata e realizzata un'architettura che implementa il modello uniforme di annotazione definito nel TR2.1.3, e che permette la gestione ed il versioning delle annotazioni, dei documenti annotati e delle ontologie utilizzate. La piattaforma sviluppata si basa su di una architettura Client/Server, in cui la parte client, che è installata sulla macchina dell'utente, consiste essenzialmente nello strumento annotatore, mentre la parte Server risiede su di un'unica macchina, accessibile da remoto, predisposta alla funzione di repository centrale. Lo strumento sviluppato (denominato preliminarmente **OverFA**) non si compone soltanto dell'annotatore, ma prevede una articolata architettura per la gestione dei contenuti e delle annotazioni. Per ulteriori dettagli si vedano il Technical Report (draft) **TR2.3.3 Realizzazione di un'architettura per l'annotazione semantica e di uno strumento Prototipale**, e gli articoli: F. Moscato, B. Di Martino, S. Venticinque, A. Martone, "OverFA: A collaborative Framework for Semantic Annotation of Documents and Web Sites", to be published in: International Journal of Web and Grid Services (IJWGS), Inderscience Press.

Beniamino Di Martino, Angelo Martone, Francesco Moscato, Salvatore Venticinque, "A versioning based framework for semantic annotation of Web documents: OVerFA", Proc. of Int. Conf. on Methods, Moels and Information Technologies for Decision Support Systems (MTISD 2008), Lecce, Italy, 18-20 sept. 2008.

Sono state esaminate le principali metodologie e tecniche per la Classificazione documentale automatizzata, con particolare riferimento alle tecniche stocastiche (in particolare basate sulla Single Value Decomposition e le tecniche di Clustering) ed alle tecniche basate su reti neurali, entrambe di sia di tipo supervised che unsupervised. Vedi Technical Report **TR2.3.4 Stato dell' arte ed Analisi delle metodologie e delle tecniche per la Classificazione ed il Clustering**

Sono stati individuati ed analizzati i principali strumenti e le piattaforme tecnologiche, commerciali ed Open Source, per la Classificazione Documentale. Tali strumenti e tecnologie sono stati classificati in base ai modelli ed alle tecniche adottate, alle funzionalità offerte, alle piattaforme di sviluppo e di esecuzione.

PROGETTO LC3	Revisione n*	0	Del	-----
--------------	--------------	---	-----	-------



Sono stati analizzati e confrontati i principali strumenti Open Source per la definizione di reti neurali tra i quali Joone e JavaNNS. Per quanto riguarda gli strumenti Open Source che implementano tecniche di Clustering si è analizzato il software TIGR MeV. Vedi Technical Report **TR2.3.5 Stato dell' arte ed Analisi delle tecnologie e degli strumenti per la Classificazione Documentale** per ulteriori dettagli.

Si sta realizzando uno strumento prototipale che permette la generazione automatica di annotazioni semantiche e per la classificazione automatica di documenti basata sull'utilizzo di reti neurali back-propagation (non supervisionate) e reti supervisionate. Vedi Technical Report **TR2.3.6 (DRAFT) Realizzazione di uno strumento prototipale per l' annotazione semantica automatica di corpora documentali basato su tecniche di classificazione automatica** per ulteriori dettagli.

Nel seguito si descrivono i risultati conseguiti dall' Unità di ricerca SPACE, coadiuvata dal Consulente Istituto Luigi Sturzo.

E' stata definita l'architettura e le specifiche funzionali che descrivono come il MAG *Repository*, che si occupa dello *storage* e della gestione dei contenuti digitali a livello centralizzato, dialoghi con le componenti software esterne ad esso, relative della produzione di metadati, all'annotazione semantica e all'IR semantico.

Sono state individuate le funzionalità fondamentali necessarie alla comunicazione tra le varie componenti previste dal progetto e la MAG - teca digitale prestando attenzione allo studio del flusso dati nella comunicazione dei moduli software formalizzata mediante *sequence diagram*. In particolare è stata definita una metodologia per accordare i dati prodotti dall'annotatore semantico a quelli attesi dallo storage centrale. Ulteriori dettagli sono reperibili nel Technical Report **TR2.3.7 Architettura generale e intercambio dei dati semantici tra il MAG Repository e lo strumento di annotazione semantica**

Il supporto fornito dal Consulente ILS, sulla base dell'esperienza maturata come consulente del Ministero per i Beni e le Attività Culturali sul progetto della Biblioteca Digitale Italiana e Network Turistico Culturale, nonché come produttori di specifiche outlines narrative, è descritto nel report **TR2.3.8 Supporto alla definizione delle specifiche funzionali del Sistema MAG Teca**

PROGETTO LC3	Revisione n*	0	Del	-----
--------------	--------------	---	-----	-------



PUBBLICAZIONI PRODOTTE

F.Moscato, B.Di Martino: "Semantic Web and Semantic Information Management", Int. J. Web and Grid Services, Vol. 4, No. 3, 2008.

F. Moscato, B. Di Martino, S. Venticinque, A. Martone, "OverFA: A collaborative Framework for Semantic Annotation of Documents and Web Sites", to be published in: International Journal of Web and Grid Services (IJWGS), Inderscience Press.

Beniamino Di Martino, Angelo Martone, Francesco Moscato, Salvatore Venticinque, "A versioning based framework for semantic annotation of Web documents: OVerFA" , Proc. of Int. Conf. on Methods, Moels and Information Technologies for Decision Support Systems (MTISD 2008), Lecce, Italy, 18-20 sept. 2008.

PROGETTO LC3	Revisione n*	0	Del	-----
--------------	--------------	---	-----	-------