

20
1996-2016

UNIVERSITÀ DELLA CALABRIA

Laboratorio di
Documentazione



ASS.I.TERM
ASSOCIAZIONE ITALIANA PER LA TERMINOLOGIA



XXVI Convegno Internazionale Ass.I.Term
Terminologia e organizzazione della conoscenza
nella conservazione della memoria digitale

Liste di autorità per l'indicizzazione e la conservazione digitale dell'archivio del CNUCE e della CGI

Maria Taverniti
Elena Cardillo

*Istituto di Informatica e Telematica, sede di Cosenza
Consiglio Nazionale delle Ricerche*

 Sede di Cosenza

Contesto



2013: Accordo di collaborazione tra Ufficio sistemi informativi e documentali del CNR e l'Istituto di Informatica e Telematica del CNR

- ✓ Sviluppare un sistema integrato per l'accesso all'informazione e la conservazione a lungo termine dei documenti del CNR
 - Digitalizzazione e indicizzazione del patrimonio storico del CNR
 - Documenti sul Centro Nazionale Universitario per il Calcolo Elettronico
 - Documenti sulla Commissione Generale Informatica

Obiettivo

Creazione di liste di autorità ad hoc, integrate e condivisibili, volte all'indicizzazione e conservazione digitale dei documenti storici del CNUCE e della CGI

Controllo d'autorità

- uso di un singolo termine per ogni [entità](#) (autore, organizzazione, ecc.)
- rinvio da forme varianti del nome a questa unica forma per consentire l'accesso a tutte le opere di un [autore](#) a prescindere da dalla forma con cui è nominato: es. *Biorci, Prof. Biorci*

Problematiche

- per aree di conoscenza di nicchia, molte organizzazioni hanno l'esigenza di creare le proprie liste di autorità localmente e i propri authority file
 - diverse forme del nome usate per la stessa persona
 - pochi riferimenti importanti
 - copertura limitata dei nomi di persone estratti dal corpus in *authority file* esterni

Approccio

Fasi principali

- a) Censimento e Digitalizzazione del patrimonio di interesse storico-scientifico
- b) Estrazione terminologica
- c) Arricchimento delle liste di autorità
- a) Mapping al Virtual International Authority File (VIAF)

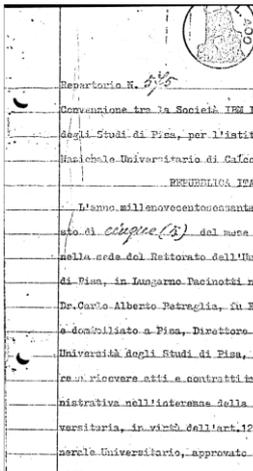
Patrimonio storico

1965: nasce il CNUCE per iniziativa di Alessandro Faedo, rettore dell'Università di Pisa, in collaborazione con IBM

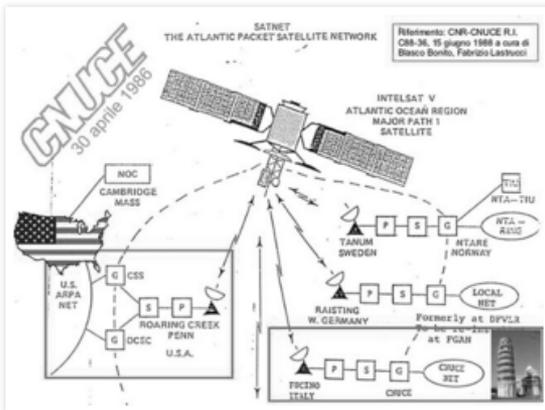
1973: il CNUCE diventa parte del Consiglio Nazionale delle Ricerche

1978: nasce la CGI, Commissione Generale per l'Informatica, col compito di controllare, incoraggiare e moderare le iniziative del CNUCE

Italian Internet Day: Il 30 Aprile si festeggia Internet in Italia



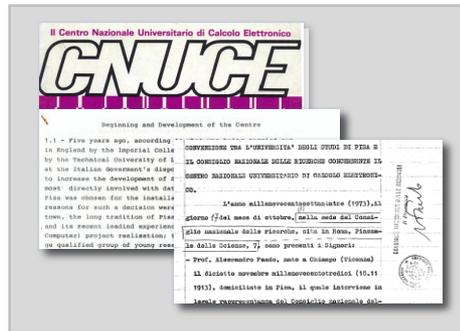
Pubblicata il: 1/04/2016



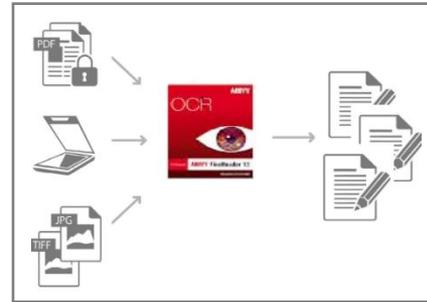
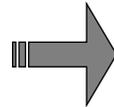
30 aprile 1986 è la data che ha cambiato l'era moderna almeno in Italia. Infatti è il giorno in cui il Cnr di Pisa si collegò ad Arpanet, l'antenato di Internet. Proprio per festeggiare i 30 anni da questo evento il presidente del consiglio Matteo Renzi ha istituito l'Italian Internet Day che appunto accadrà a

ione
 o sul documento
 gono tutti i pre-
 vima di procedere
 i lavori della Com
 precisione quali
 alla Commissione.
 esse un organismo
 Comitati Naziona
 ita secondo l'im-
 lo alla Commissio
 NUCE" con una du-
 linee di interven
 erebbero troppo
 e rilevato da
 rendendo le mosse
 ioritari, svilup-
 a tali problemi,

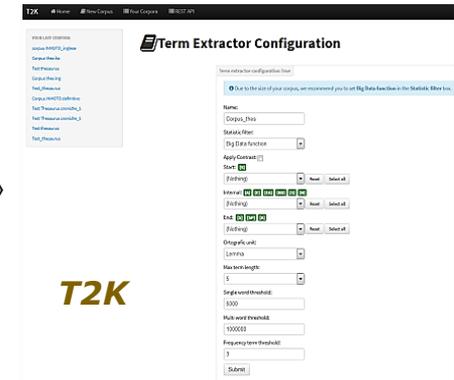
Corpus ed estrazione terminologica



**Corpus: ~ 2500 pagine
~ 360 PDF**



OCR



Estrazione terminologica

Text2Knowledge (T2K)

- Glossario di termini rilevanti usato come base per la creazione di una lista di soggetti e parole chiave
 - 1,250 termini
- Lista di Named Entities (Persone e Organizzazioni) usata come base per la creazione delle liste di autorità:
 - 530 nomi

Problematiche

- Termini rilevanti estratti troppo generici e non di dominio tecnico-scientifico
- Mancato riconoscimento di nomi di autorità (persone)

Necessaria revisione e arricchimento manuale delle liste estratte



Risultati

- **Statistiche liste di Nomi di autorità**

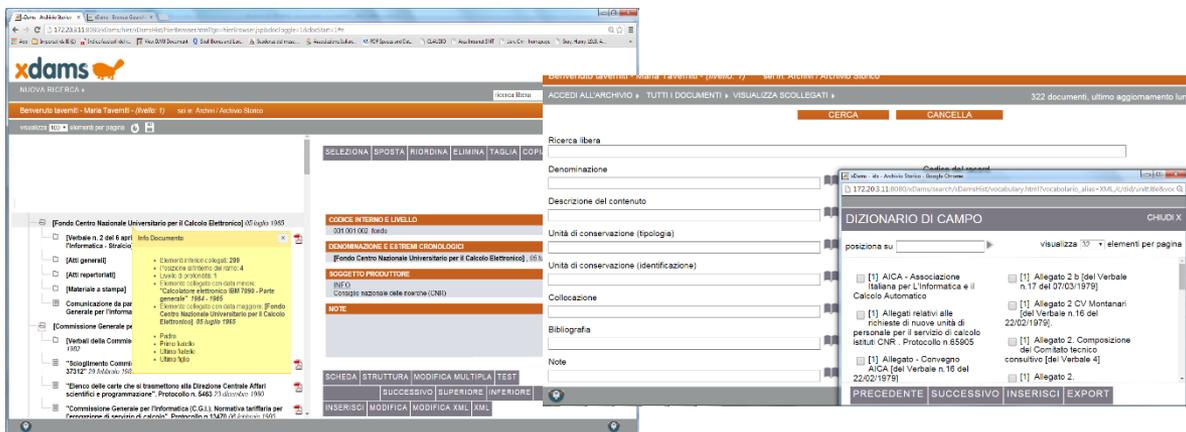
	Estrazione con T2K	Revisione Manuale	Mappatura a VIAF	Dati da Wikipedia	Dati da altre fonti
<i>Lista autorità Persone</i>	237	700	<u>420</u>	280	438
<i>Lista autorità Organizzazioni</i>	295	400	<u>110</u>	205	195
Total	532	1,100	530	485	633

- **Statistiche lista Soggetti**

	Estrazione con T2K	Revisione Manuale
<i>Lista Soggetti</i>	1.250	850

- *In corso la mappatura dei candidati a Soggetto al Nuovo Soggettario e LCSH*

Descrizione documentale e integrazione nella S&TDL



Per rendere la documentazione ricercabile, fruibile e dialogabile con la piattaforma S&TDL è stato effettuato un processo di:

descrizione attraverso linguaggi, metodi e strumenti conformi

- agli standard ISAD, ISAAR (CPF)
- allo standard Dublin Core
- al modello dati basato su elementi dello standard EAD

Integrazione nella piattaforma xDams delle liste d'Autorità del CNUCE/CGI per l'indicizzazione dei documenti

integrazione nella piattaforma S&TDL attraverso l'export in XML dell'archivio

Conclusioni

La creazione e l'uso di liste di autorità ad hoc estratte dai documenti CNUCE/CGI permette

- un controllo delle varianti dei nomi
- una più efficiente disambiguazione degli stessi
- un accesso mirato ai rispettivi documenti nella *Digital Library*

Benefici

- riutilizzo di una lista di autorità consolidata
- identificazione di persone e dei *topics* a loro connessi
- record di nomi standardizzati e allineati a VIAF
 - liste di autorità internazionali
 - ricerca in diverse lingue attraverso identificativi interconnessi a livello globale

Work in progress

Rappresentazione delle liste d'autorità del CNUCE/CGI in RDF in Linked Open Data



Maria Taverniti maria.taverniti@cnr.it

Elena Cardillo elena.cardillo@iit.cnr.it