



Digital Humanities



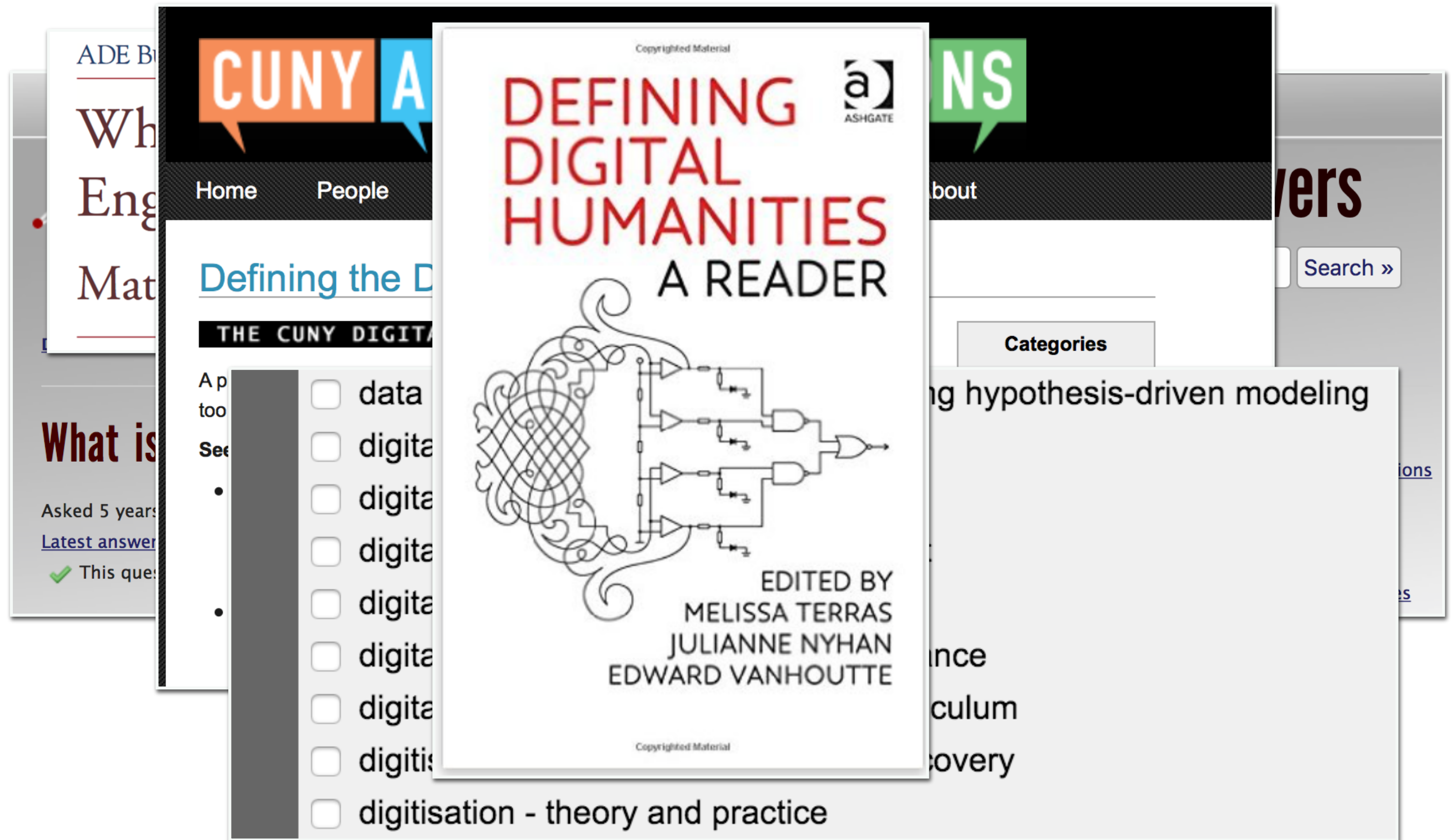
Sara Tonelli

Giugno 2019

Digital Humanities

Sara Tonelli
satonelli@fbk.eu

Cosa sono le Digital Humanities?



Una definizione

“Digital Humanities interprets the cultural and social impact of new media and information technologies, as well as creates and applies these technologies to interrogate cultural, social, historical, and philological questions.” (*UCLA DH Center*)

Manifesto delle Digital Humanities

[...] For us, the digital humanities concern the totality of the **social sciences and humanities**. The digital humanities are not *tabula rasa*. On the contrary, they rely on all the paradigms, *savoir-faire* and knowledge specific to these disciplines, while mobilizing the tools and unique perspectives enabled by digital technology.

The digital humanities designate a “**transdiscipline**”, embodying all the methods, systems and heuristic perspectives linked to the digital within the fields of humanities and the social sciences.[...]

CONTEXT
We, professionals or observers of the digital humanities (humanités numériques) came together in Paris for THATCamp on May 18th and 19th, 2010. Over the course of these two days, we discussed, exchanged, and collectively reflected upon what the digital humanities are, and tried to imagine and invent what they could become. At the close of the camp – which represents but a first step – we propose to the research communities, and to all those involved in the creation, publication, valorization or preservation of knowledge, a manifesto for the digital humanities.

DEFINITION
1. Society's digital turn changes and calls into question the conditions of knowledge production and distribution.
2. For us, the digital humanities concern the totality of the social sciences and humanities. The digital humanities are not *tabula rasa*.

On the contrary, they rely on all the paradigms, *savoir-faire* and knowledge specific to these disciplines, while mobilizing the tools and unique perspectives enabled by digital technology.
3. The digital humanities designate a “transdiscipline”, embodying all the methods, systems and heuristic perspectives linked to the digital within the fields of humanities and the social sciences.

SITUATION
4. We observe:
• that experiments in the digital domain of the social sciences and humanities have multiplied in the last half century. What have emerged most recently are centers for digital humanities – which at the moment are themselves only prototypes or areas of application specific to the approach of digital humanities;
• that computational and digital approaches have greater technical, and therefore economic, research constraints; that these constraints provide an opportunity to foster collaborative work;
• that while a certain number of proven methods exist, they are not equally known or shared;

• that there are many communities deriving from shared interests in practices, tools, and various interdisciplinary goals – encoding textual sources, geographic information systems, lexicometry, digitization of cultural, scientific and technical heritage, web cartography, datamining, 3D, oral archives, digital arts and hypermedia literatures, etc. – and that these communities are converging to form the field of digital humanities.

DECLARATION
5. We, professionals of the digital humanities, are building a community of practice that is solidary, open, welcoming and freely accessible.
6. We are a community without borders. We are a multilingual and multidisciplinary community.
7. Our objectives are the advancement of knowledge, the improvement of research quality in our disciplines, the enrichment of knowledge and of collective patrimony, in the academic sphere and beyond it.
8. We call for the integration of digital culture in the definition of the general culture of the twenty-first century.

MANIFESTO
FOR THE DIGITAL HUMANITIES

GUIDELINES
9. We call for open access to data and metadata, which must be documented and interoperable, both technically and conceptually.
10. We support the dissemination, exchange and free modification of methods, code, formats and research findings.
11. We call for the integration of digital humanities education within social science and humanities curricula. We also wish to see the creation of diplomas specific to the digital humanities, and the development of dedicated professional education. Finally, we want such expertise to be considered in recruitment and career development.
12. We commit to building a collective expertise based upon a common vocabulary, a collective expertise proceeding from the work of all the actors involved. This collective expertise is to become a common good. It is a scientific opportunity, but also an opportunity for professional insertion in all sectors.
13. We want to help define and propagate best practices, corresponding to needs identified within or across disciplines, which should derive and evolve from debate and consensus within the communities concerned. The fundamental openness of the digital humanities nevertheless assures a pragmatic approach to protocols and visions, which maintains the right to coexistence of different and competing methods, to the benefit of both thought and practice.
14. We call for the creation of scalable digital infrastructures responding to real needs. These digital infrastructures will be built iteratively, based upon methods and approaches that prove successful in research communities.

FOR FURTHER INFORMATION
Web : <http://tcp.hypotheses.org>
Email : thatcamp@paris-revues.org
Twitter : <http://twitter.com/thatcampparis>
Discussion list : dh@cru.fr

JOIN US

YOU CAN SIGN THE MANIFESTO ONLINE

THATCamp is a trademark of the Center for Writing and New Media
THATCamp Paris was initiated by the Center for Open Electronic Publishing (Chen) and was jointly organized by seven Paris-based partners. <http://dh.cru.fr>

Cosa ci motiva

- I problemi nel mondo “reale” sono **interdisciplinari**
- Ci sono domande importanti che trovano risposte adeguate soltanto nell’interazione tra discipline diverse
- Problemi nuovi legati alla **complessità del mondo contemporaneo** hanno bisogno di risposte nuove
- Ma anche problemi vecchi possono trovare **soluzioni nuove!**

Perché Digital Humanities?

- Fornire ampio **accesso alle informazioni** culturali collezioni, archivi ed edizioni digitali
- **Manipolare dati** per analizzarli in maniera innovativa, gestire, modellare, estrarre, modellare informazioni
- Trasformare la **divulgazione** dalla carta al digitale
- **Cambiare l'insegnamento** e l'apprendimento, nuovi strumenti da usare in classe per sviluppare nuove competenze
- Avere un **impatto** pubblico, coinvolgere il pubblico con progetti collaborativi

Informatica Umanistica

Digital Humanities = Informatica Umanistica

Scienze umane + Scienze esatte + Scienze Applicate

Quando è nata l'Informatica Umanistica

- 1949: Padre Roberto Busa incontra il presidente di IBM Thomas Watson per chiedere aiuto nell'analisi del lessico di Tommaso d'Aquino



Lemma 06079
anima: anima animae
Type of lemma: (A) common word
Type of meaning: (NV) invisible thing

cases	c. freq.	places	pl. freq.	form	type	infl.	num.	gen.	case	deg.	voice	tense	mood	pers.	comp.	notes
12983	1,22%	6780	7,78%	☒ anima	n.	irr. decl. sing. f.			nom.	pos.						b
8511	0,80%	4921	5,65%	☐ animae	n.	irr. decl. sing. f.			gen.	pos.						g
60	0,01%	26	0,03%	☐ anime	n.	irr. decl. sing. f.			gen.	pos.						h
1	0,00%	1	0,00%	☐ animaeque	n.	irr. decl. sing. f.			gen.	pos.					-que	
4325	0,41%	2694	3,09%	☐ animam	n.	irr. decl. sing. f.			acc.	pos.						
1	0,00%	1	0,00%	☐ animamque	n.	irr. decl. sing. f.			acc.	pos.					-que	
609	0,06%	419	0,48%	☐ animarum	n.	irr. decl. pl. f.			gen.	pos.						
349	0,03%	278	0,32%	☐ animabus	n.	irr. decl. pl. f.			dat.	pos.						
743	0,07%	547	0,63%	☐ animas	n.	irr. decl. pl. f.			acc.	pos.						b

Analisi lessicografica di 9 milioni di parole
Dalle schede perforate ai nastri magnetici fino ai CD e al web

Dove si studia l'Informatica Umanistica



Associazione AIUCD

ASSOCIAZIONE per
l'INFORMATICA UMANISTICA
e la CULTURA DIGITALE

[HOME](#)[CHI SIAMO](#)[PUBBLICAZIONI](#) ▾[PROGETTI](#)[DIDATTICA](#)[CENTRI](#)[CONVEGNI](#)[NEWS](#)

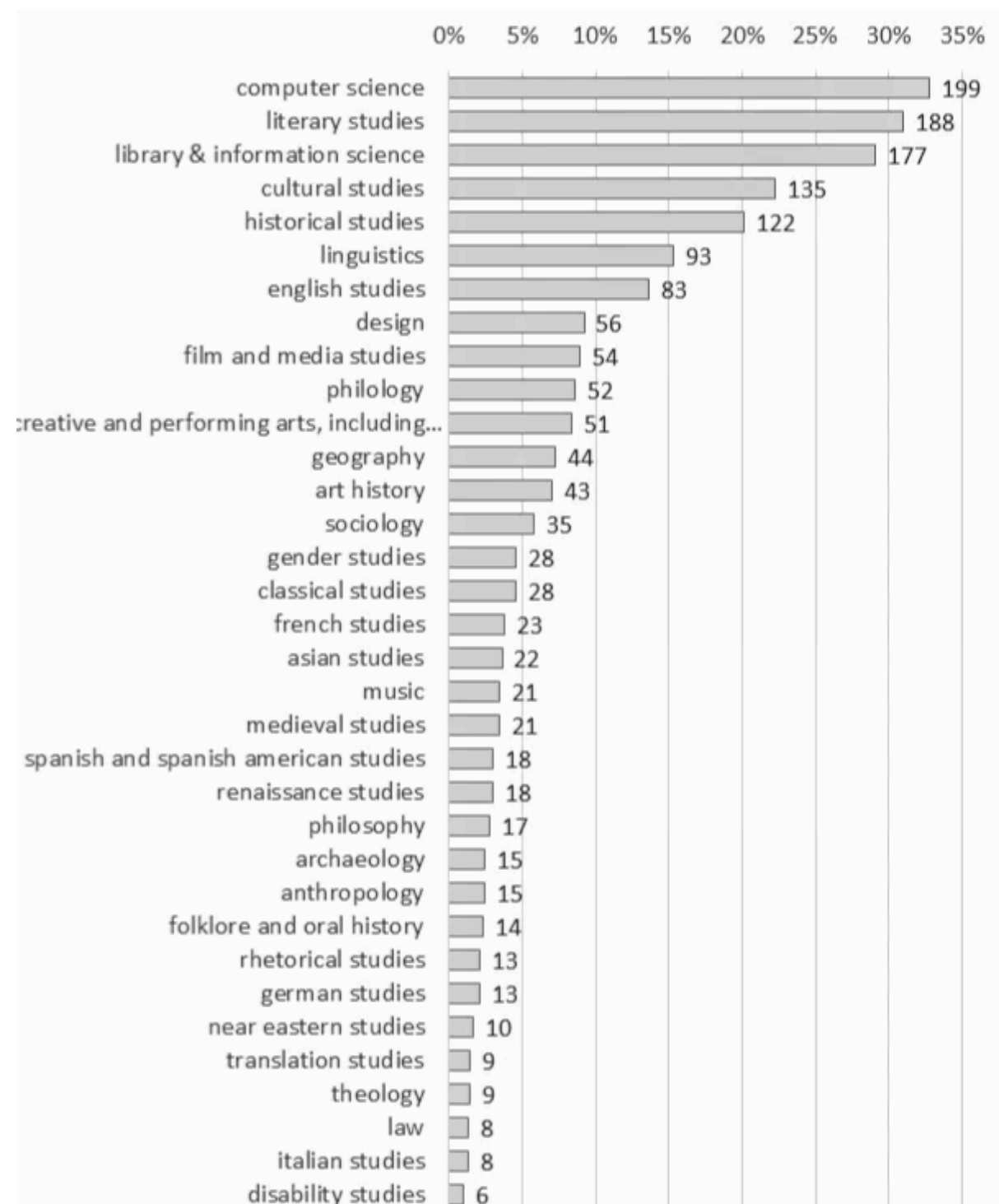
ASSOCIAZIONE PER L'INFORMATICA UMANISTICA E LA CULTURA DIGITALE

AIUCD diffonde la riflessione metodologica e teorica, la collaborazione scientifica e lo sviluppo di pratiche, risorse e strumenti condivisi nel campo dell'informatica umanistica e nell'uso delle applicazioni digitali in tutte le aree delle scienze umane. Promuove inoltre la riflessione sui fondamenti umanistici delle metodologie informatiche nel campo delle culture di rete.

AIUCD fosters methodological and theoretical research, scientific collaboration and the development of shared practices, resources and tools in the Digital Humanities, as well as digital applications in all Humanities disciplines. Additionally, it seeks to stimulate reflection on the theoretical foundations of computational methods in Internet culture.

[CHI SIAMO](#)[DIVENTA SOCIO](#)

Quali tematiche?



Articoli scientifici presentati
alla conferenza Digital
Humanities 2017

Archivi digitali

- Necessità di raccogliere, preservare e presentare il passato via web
- La digitalizzazione ci permette di sfruttare caratteristiche come “*capacity, accessibility, flexibility, diversity, manipulability, interactivity and hypertextuality*” (Cohen and Rosenzweig, 2005).
- Alcuni potenziali problemi sono legati invece al mezzo digitale e non a quello cartaceo: “*quality, durability, readability, passivity and inaccessibility*” (Cohen and Rosenzweig, 2005).

Progetto Epistolario De Gasperi

EDIZIONE
NAZIONALE
EPISTOLARIO
ALCIDE
DE GASPERI

[IT](#) / [EN](#)



[L'EDIZIONE NAZIONALE](#) [ALCIDE DE GASPERI](#) [L'EPISTOLARIO](#) [COLLABORA AL PROGETTO](#)

[Progetto](#)

[Commissione Scientifica](#)

[Network](#)

[Contatti](#)



Edizione Nazionale dell' Epistolario
di Alcide De Gasperi

<http://www.epistolariodegasperi.it/>

Architettura di trascrizione

Tre profili



Trascrittore



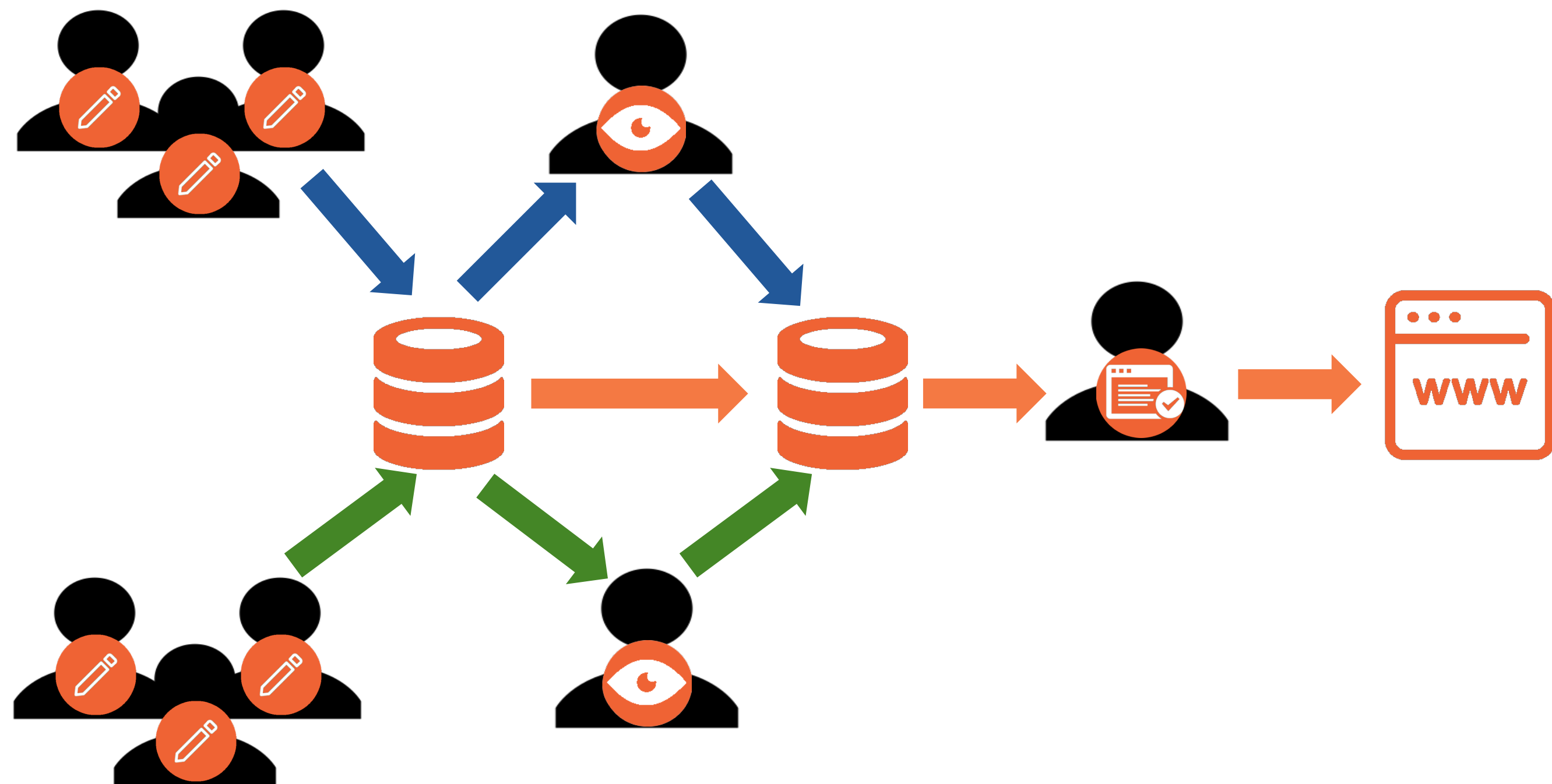
Supervisore



Editore



Trascrizione e pubblicazione online



Tool di Trascrizione

Titolo

 Da Alcide De Gasperi A Giovanni Ciccolini , 14/02/1925 

Mittente/i

Nome

 Alcide

Cognome

 De Gasperi

Genere: ☒ M ☐ F

Dato provvisorio: ☐

Luogo

Trento, Venezia tridentina, Italia 

Dato provvisorio: ☐

Destinatario/i

Nome

 Giovanni

Cognome - Destinatario collettivo/altra denominazione

 Ciccolini

Genere: ☒ M ☐ F

Dato provvisorio: ☐

Luogo

NA 

Dato provvisorio: ☐

Data del documento

 14/02/1925

Dato provvisorio: ☐

Data di acquisizione

 11/06/2018

Tipologia

Lettera 

Tema

 Tema
Vita Privata
Politica Nazionale
Politica Internazionale
Politica Locale

Tipo di scritto

Personale 

Tool di Trascrizione

Esemplare	Diritti di utilizzo
Originale	Dominio pubblico / nessuna restrizione d'uso

Lingua	Busta	Descrizione Busta
Italian	☐	

Carta Intestata	Intestazione	Numero Carte	Larghezza	Altezza	Tecnica di scrittura
☐	 Intestazione	1	210 mm	297 mm	Manoscritta

Descrizione fisica	Firma autografa	Firma
 Lettera vergata su carta di colore bianco, con inchiostro di colore bruno scuro	☒	 Degasperi

Stato di conservazione
Buono

Deposito/Archivio	Sigla	Segnatura
 Deposito/Archivio	Codice Arch	 Segnatura

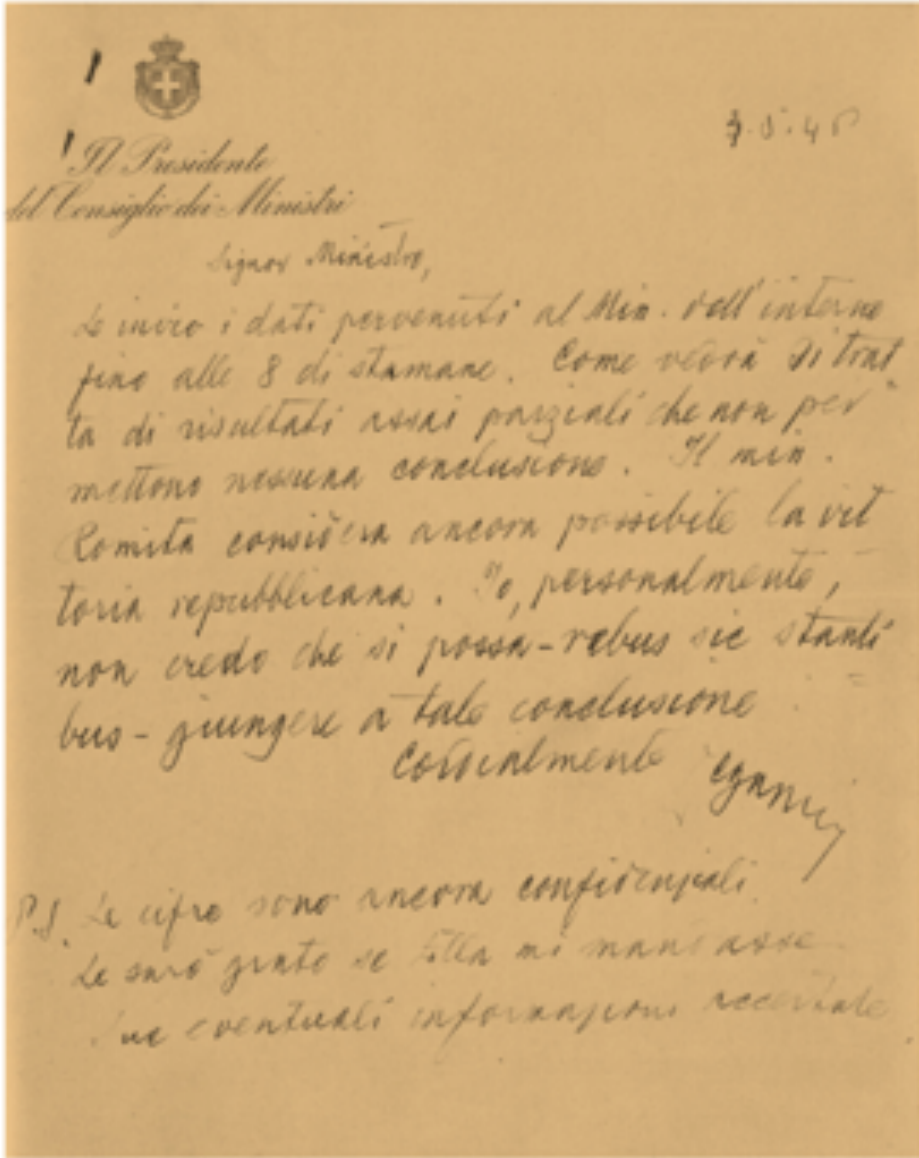
Edita
 Edita

Commento sui metadati

Tool di Trascrizione



Documenti



Trascrizione:

4.6.46

Signor Ministro,
Le invio i dati pervenuti al Min. dell'interno fino alle 8 di stamane. Come vedrà si tratta di risultati assai parziali che non permettono nessuna conclusione. Il min. Romita ritiene ancora possibile la vittoria repubblicana. Io, personalmente, non credo che si possa - rebus sic stantibus - giungere a tale conclusione.

Cordialmente Degasperì

P.S. Le cifre sono ancora confidenziali. Le sarò grato se Ella mi mandasse Sue eventuali informazioni accertate

Postille

Ab No S C I X

[DEMO]

Database di lettere

<https://www.epistolariodegasperi.it>

Trascrizione automatizzata

Transcribe. Collaborate.
Share...

...and benefit from cutting edge research in
Handwritten Text Recognition!

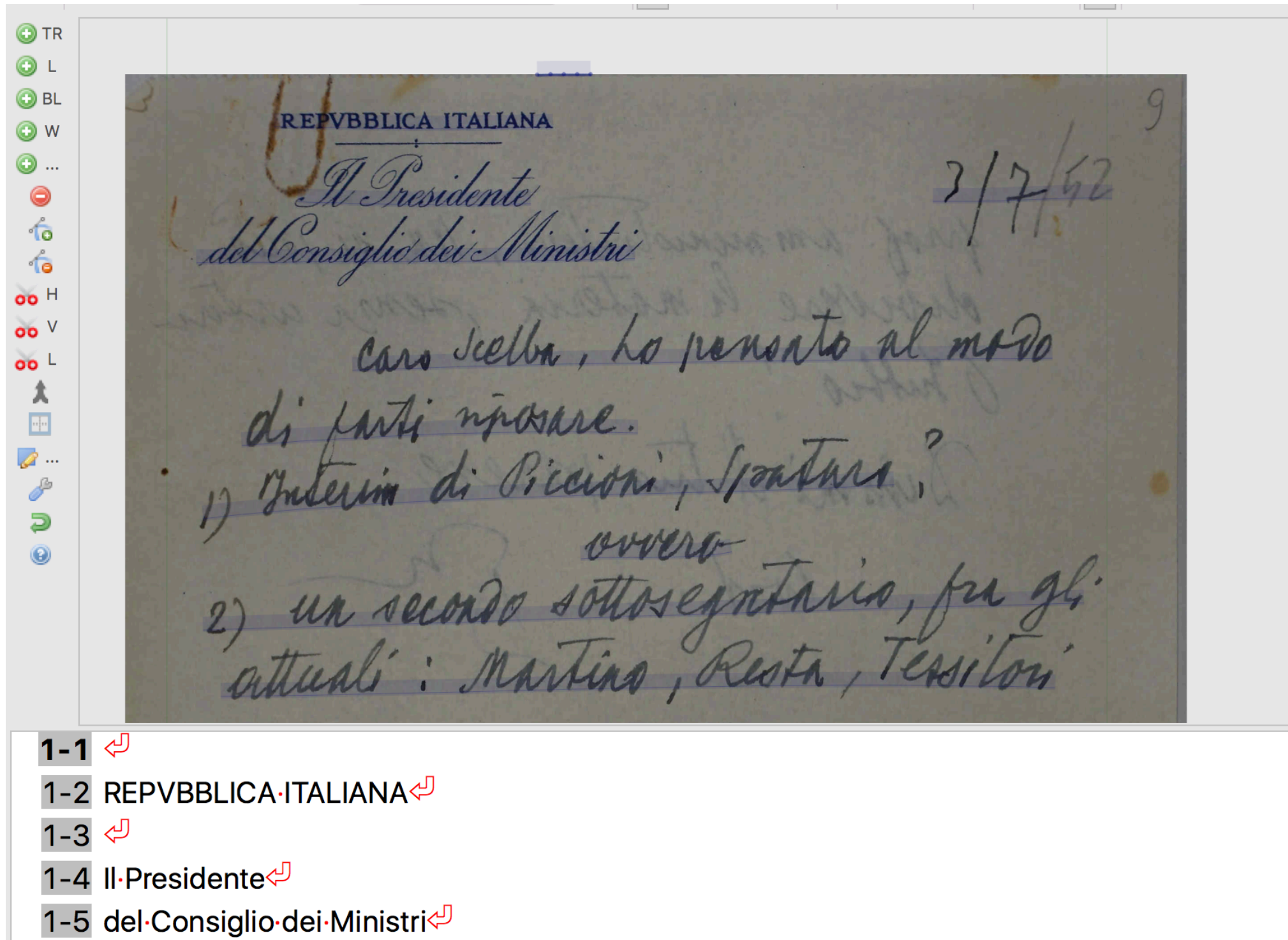
[Download version 1.6](#)

[Download version 1.6 for Mac](#)

[Wiki » How-to guide \(pdf\) »](#)

<http://transkribus.eu>

Trascrizione automatizzata

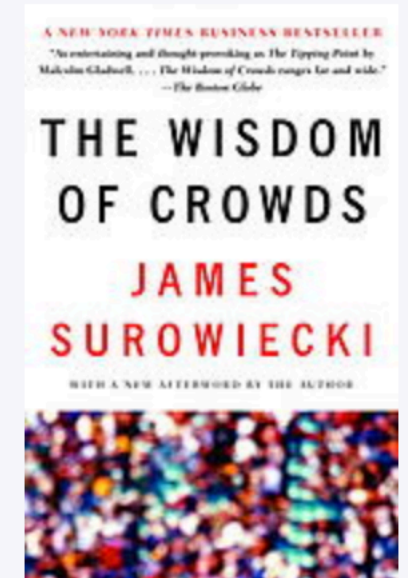


Progetti collaborativi

Saggezza della folla: teoria sociologica secondo cui un insieme di individui inesperti sarebbe comunque in grado di fornire una risposta adeguata e valida a una domanda, più di quanto non siano in grado di farlo degli esperti

Nelle Digital Humanities e in linguistica, diffuso utilizzo del **crowdsourcing** (es. **Amazon Mechanical Turk**)

Esempi di successo: Bentham Project, con le attività relative a **Transcribe Bentham**: 90.000 immagini, 20.000 pagine trascritte



Cover of mass market edition by Anchor

Author	James Surowiecki
Country	United States
Language	English
Publisher	Doubleday; Anchor
Publication date	2004
Pages	336
ISBN	978-0-385-50386-0
OCLC	61254310 
Dewey Decimal	303.3/8 22
LC Class	JC328.2 .S87 2005

Crowdsourcing

“Crowdsourcing is a type of participative online activity in which an individual, an institution, a non-profit organization, or company proposes to a group of individuals of varying knowledge, heterogeneity, and number, via a flexible open call, the voluntary undertaking of a task”

Estellés-Arolas and González-Ladrón-De-Guevara (2012)

Vantaggi:

- Creazione di una **comunità** di interessati che ruota intorno a un progetto
- **Risparmio** in termini di tempo e risorse economiche senza perdere qualità
- **Maggiore** motivazione, confronto, (gamification)

Crowdsourcing

Svantaggi / Rischi: Un esempio

- Annotare se l'attitudine di De Gasperi verso il *sindacato* era positiva o negativa
- Annotazione di 525 frasi contenenti la parola *sindacato* / *sindacalismo* annotate da 2 esperti e tramite crowdsourcing (5 annotazioni) sulla piattaforma **CrowdFlower**

Crowdsourcing

Contadini , fate un viso sorridente ! Il partito dei contadini è scomparso , il partito socialriformista è scomparso , i **sindacati** non esistono . Che resta ?

In che maniera si parla della parola evidenziata?

- ☐ Positiva
- ☐ Negativa
- ☐ Neutra
- ☐ Non so

Annotazione	Accordo (Kleiss' kappa)
Accordo tra Esperto1 e Esperto2	0.45
Accordo tra Esperto1 e "crowd"	0.35
Accordo tra Esperto2 e "crowd"	0.50

Problema: come gestire il disaccordo, problemi con la definizione del task?

Visualizzazione “da lontano”

La letteratura vista da lontano, Franco Moretti, 2005, Einaudi editore.

- Sostituire alla lettura ravvicinata del testo letterario (*close reading*) una lettura “da lontano”, costruendo grafi, carte e alberi partendo da un’analisi quantitativa dei testi letterari
- Processo di deliberata **riduzione** e **astrazione**, allontanamento rispetto alla concretezza di un singolo testo
- **Lettura a distanza** come modalità di analisi che riduce i dettagli ma fa emergere forme, regolarità e relazioni tra i testi

Close Reading

- Metodo tradizionale di analisi del testo e critica letteraria
- Interpretazione basata sul rapporto ravvicinato tra lettore o critico e testo

*Unveil words, verbal images,
elements of style, sentences,
argument patterns
(Jasinski, 2001)*

Distant Reading

- Nuovo metodo per l'analisi del testo e la critica letteraria
- Interpretazione basata su modelli astratti e informazioni statistiche

*A condition of knowledge: it allows
you to focus on units that are much
smaller or much larger than the text.
(Moretti, 2000)*

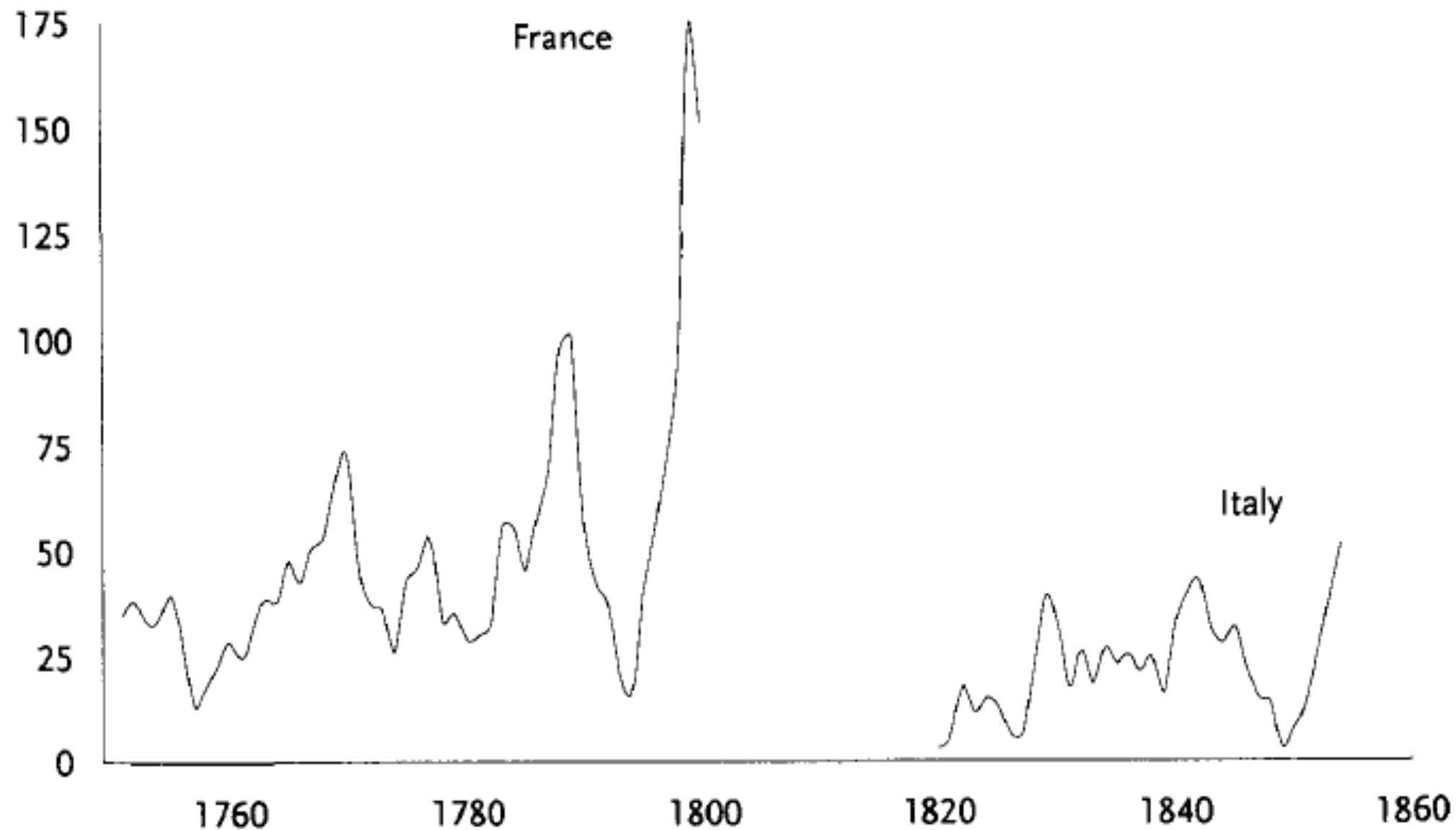
Dai testi ai modelli

Graphs, Maps, Trees: Abstract Models for a Literary History,
Franco Moretti (2007)

- Il testo viene sottoposto a una riduzione e astrazione prendendo a prestito nozioni da altre discipline:
 1. Grafi
 2. Mappe
 3. Alberi

Attraverso grafi, mappe e alberi le discipline umanistiche si dispiegano davanti ai nostri occhi sotto una veste totalmente nuova

Grafici: Diffusione del romanzo

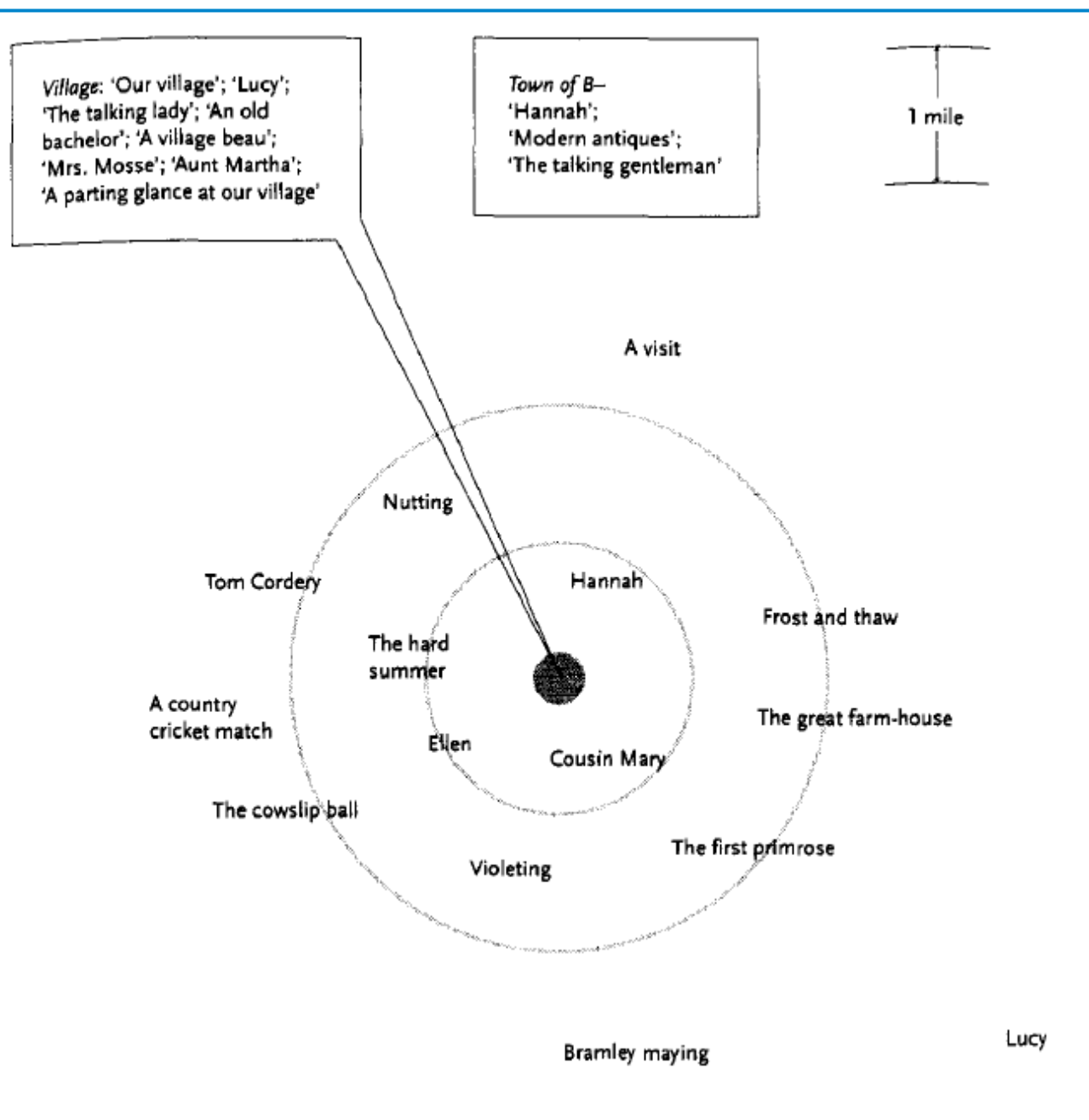


(Moretti, 2005)

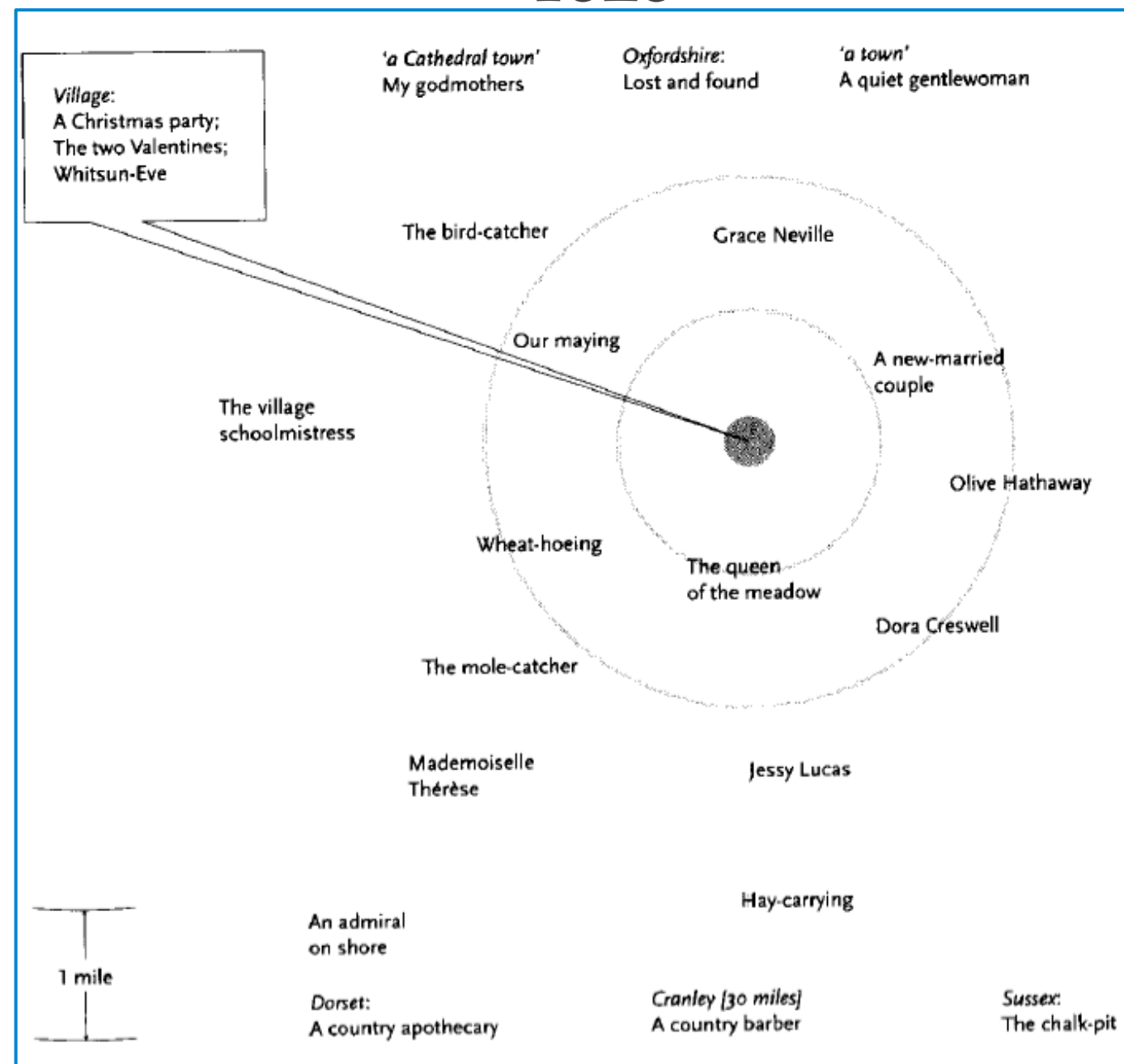
Mappe: Relazioni tra persone

Mary Mitford, "Our Village"

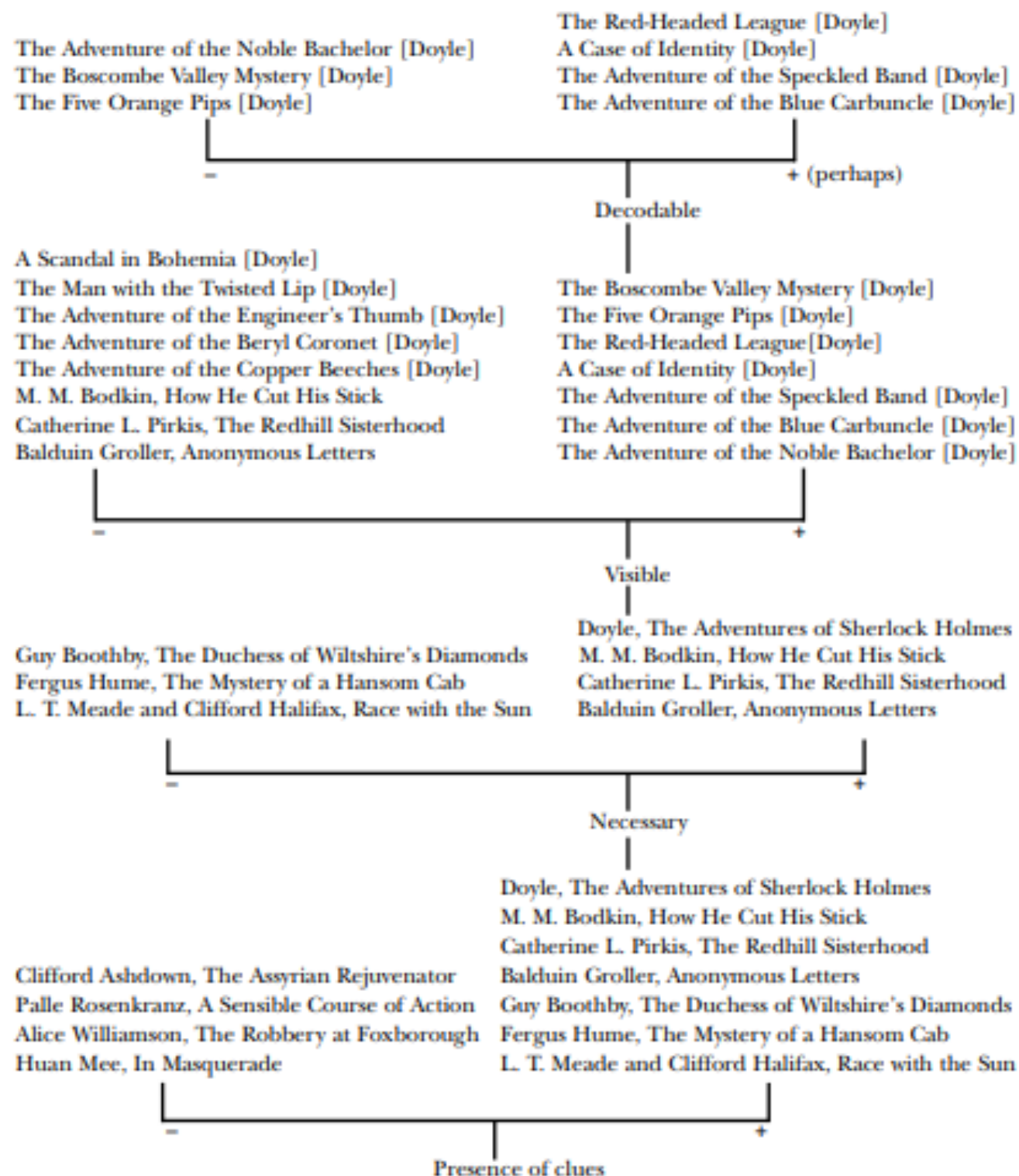
1824



1828



Alberi: Relazioni tra libri

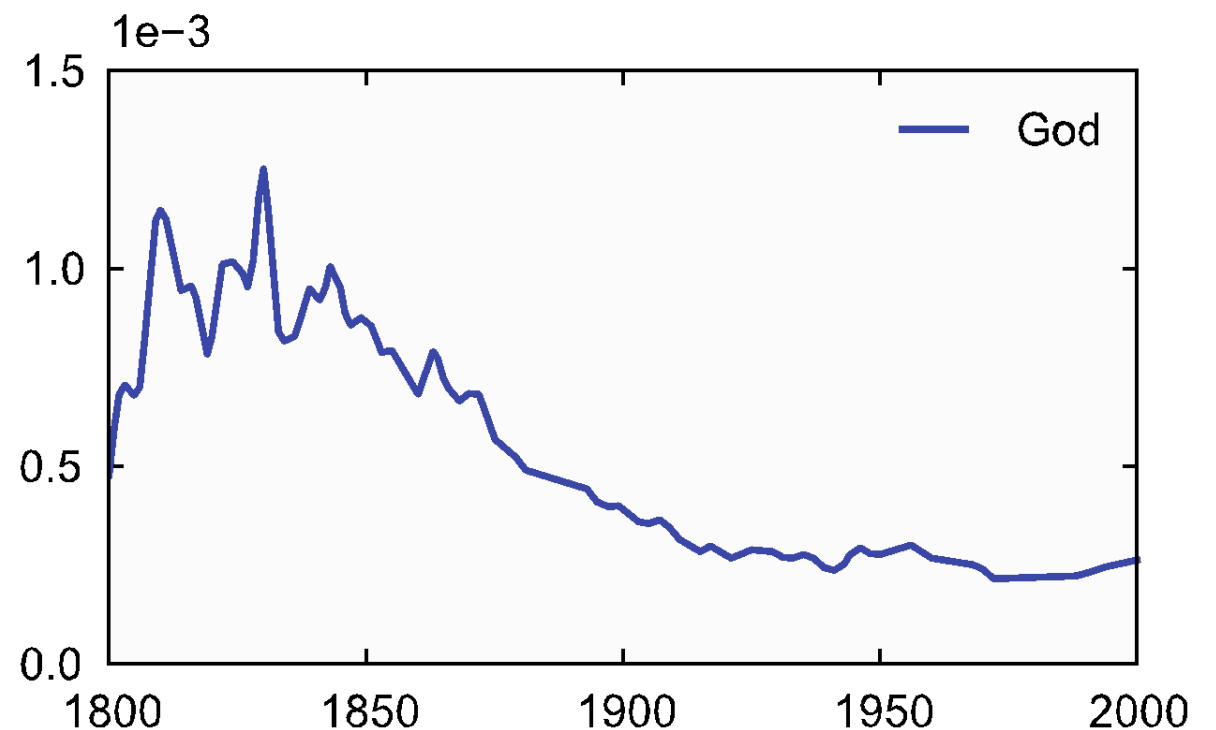
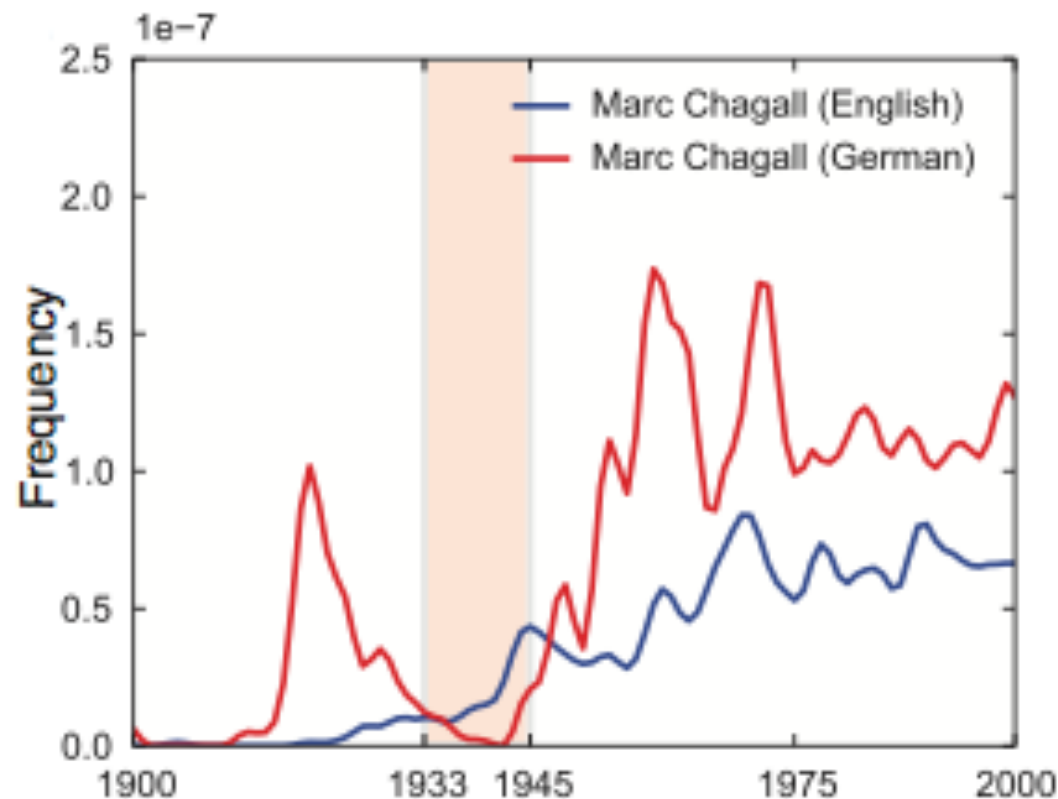


Romanzi di Conan Doyle:
il ruolo degli indizi nello
svelare il colpevole

Culturomics

Quantitative Analysis of Culture Using Millions of Digitized Books, Jean-Baptiste Michel et al., Science 331 (2010)

- *Culturomics*: application of data collection and analysis techniques to the study of human culture



<https://books.google.com/ngrams>

Prima del “distant reading”

“The great growth of statistical research in our times has made felt the need to record the results in forms less dry, more useful, and able to be explored more rapidly than numbers alone; thus, diverse representations have been imagined, among others my graphic tables and my figurative maps.”

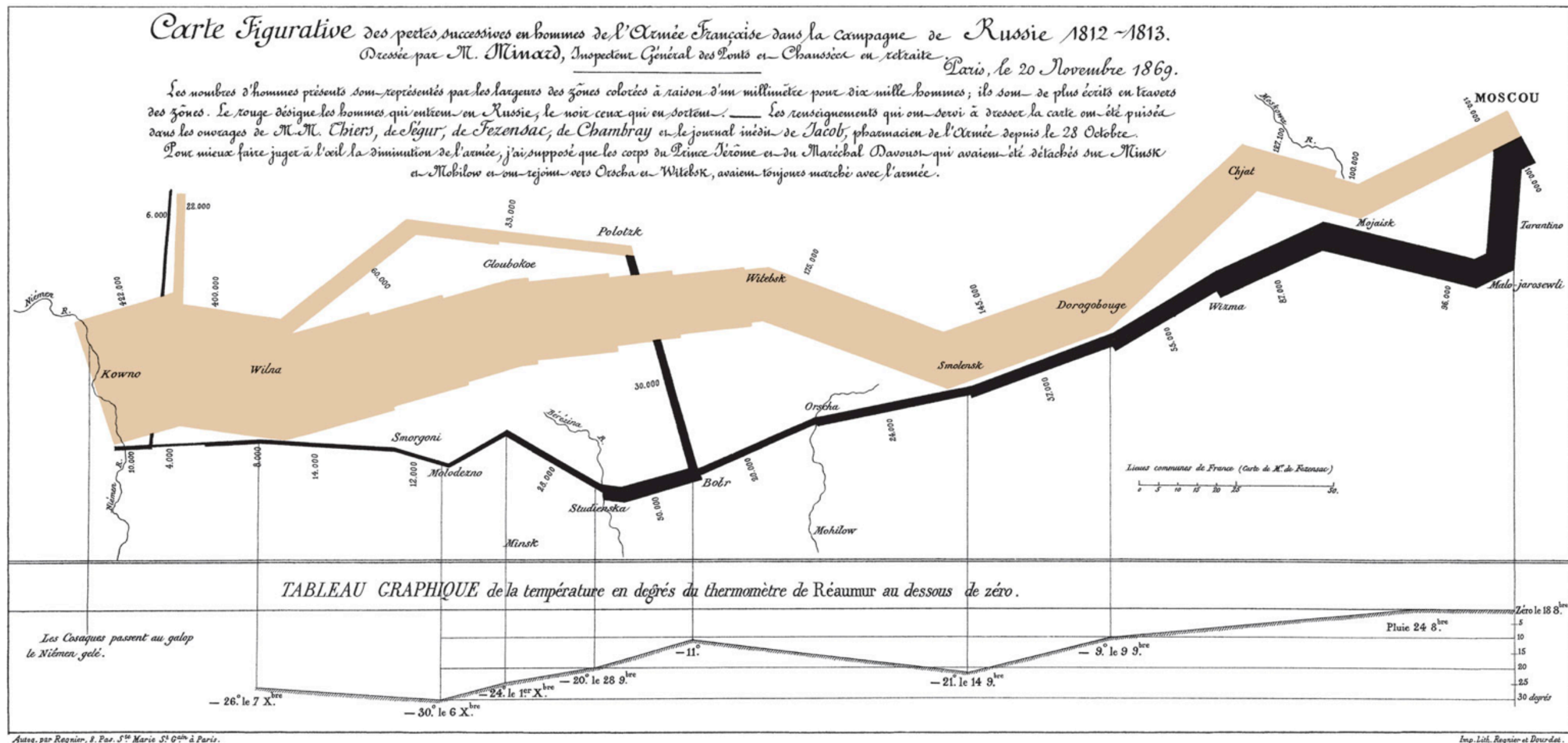
Minard, “Graphic Tables and Figurative Maps” (1862)

Importanza delle visualizzazioni!

Prima del “distant reading”

Napoleon's Russian Campaign (1861)

“May well be the best statistical graphic ever drawn”, Tufte



I passi del “distant reading”

- Raccolta di dati rilevanti
 - formato?
 - accessibilità?
 - copyright?
- Formattazione e pulizia dei dati
- Analisi dei dati / estrazione di informazioni
- Formattazione dei risultati
- Creazione della visualizzazione
- Interpretazione

Perché il “distant reading” è importante

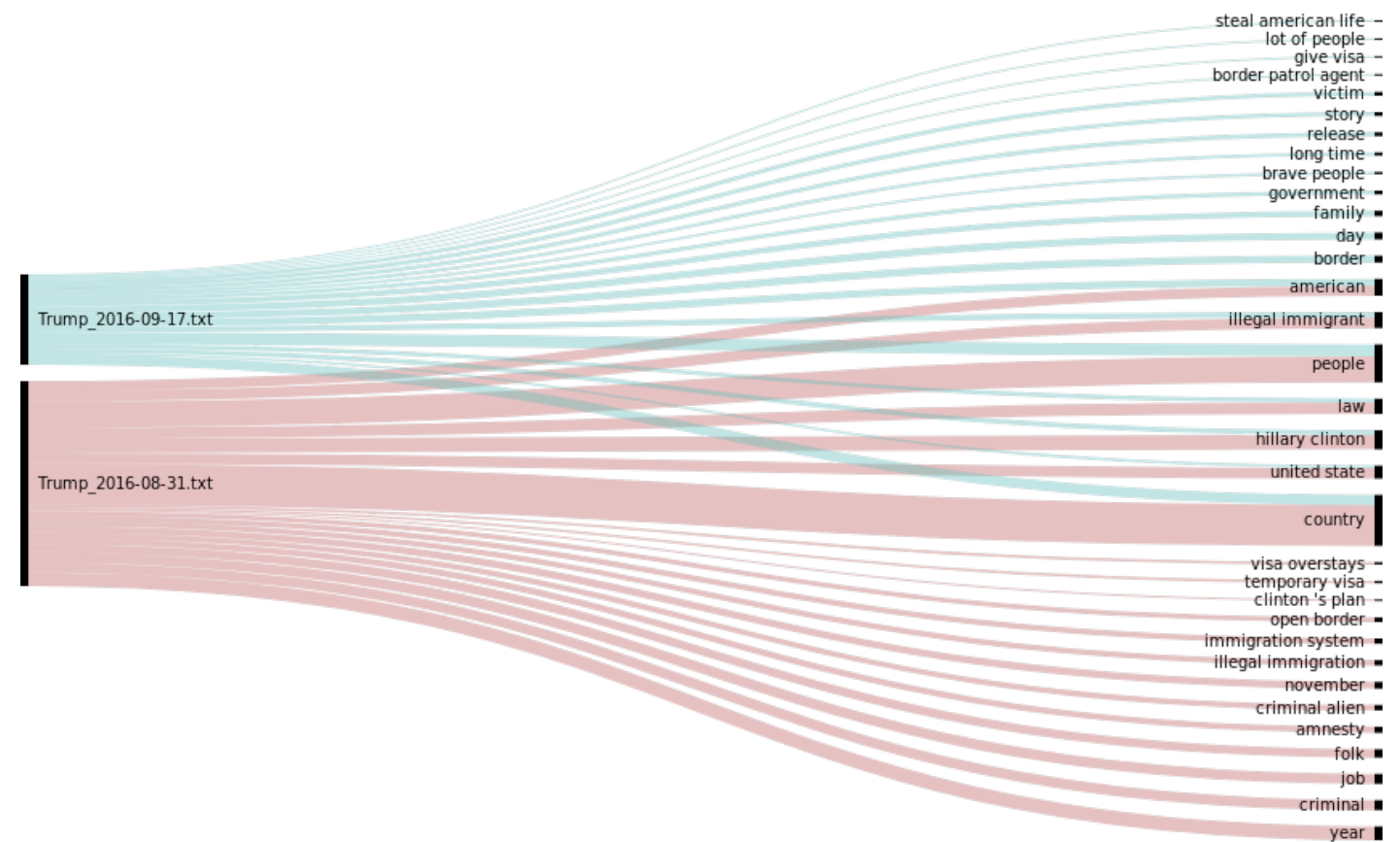
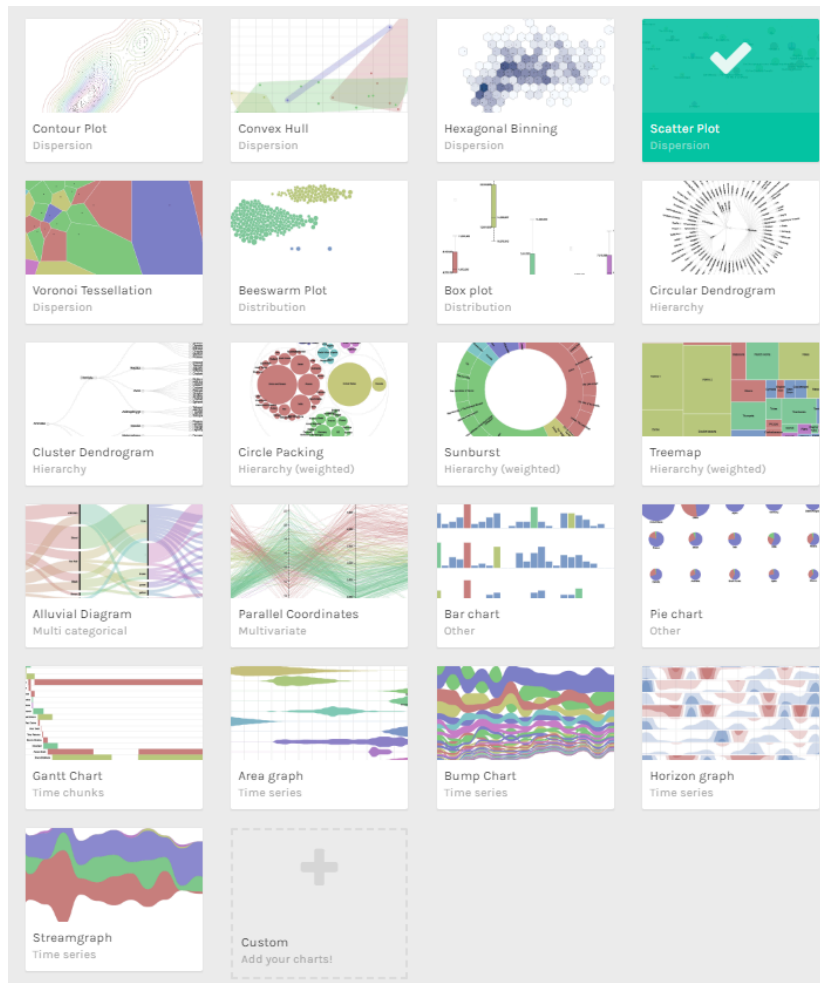
- Riproducibile e ripetibile
- Fa emergere delle regolarità che resterebbero nascoste
- Aiuta a gestire l’aumento vertiginoso di fonti digitali

MA...

Il “close reading” (analisi qualitativa) resta fondamentale e complementare al “distant reading”



Alcuni tool per il “distant reading”



<http://app.rawgraphs.io/>

Alcuni tool per il “distant reading”



<https://voyant-tools.org/>

Alcuni tool per il “distant reading”

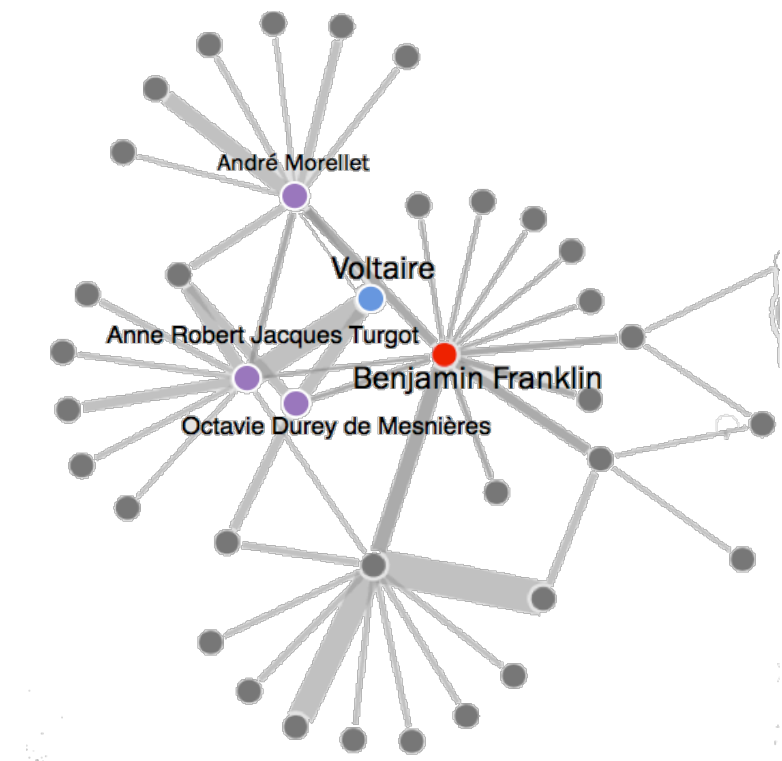
Gephi

<https://gephi.org/>



Martin Grandjean

“Mapping Shakespeare’s tragedies” - 2015



*Mapping the Republic of
Letters*

Progetto LOD Navigator



La famiglia Coen

CERCA SU TUTTA LA CDEC DIGITAL LIBRARY



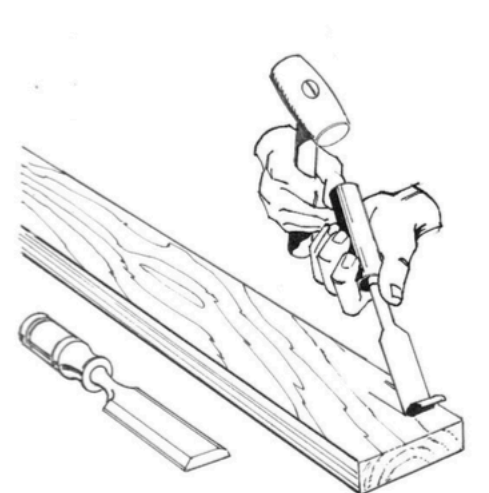
Progetto LOD Navigator

- **Motivazione:** “L’Olocausto è un fenomeno profondamente geografico” (Knowles et al., 2014)
- Dati disponibili come **Linked Open Data** non utilizzabili facilmente
- Sviluppare **visualizzazioni** che generino nuova conoscenza sulle politiche naziste e ricostruiscano storie individuali o di gruppo

Fasi del progetto

Quattro fasi:

- Osservazione dei dati
- Estrazione di informazioni
- Integrazione e correzione delle informazioni
- Visualizzazione



Osservazione dei dati

CD
EC

DIGITAL
LIBRARY

HOME

FOT

Levi, Primo

39
RDF

5

INFORMAZIONI BI

PROFESSIONE

PERSECUZIONE

Levi, Primo

<http://dati.cdec.it/lod/shoah/person/5002>

ENTITÀ DI TIPO: Person

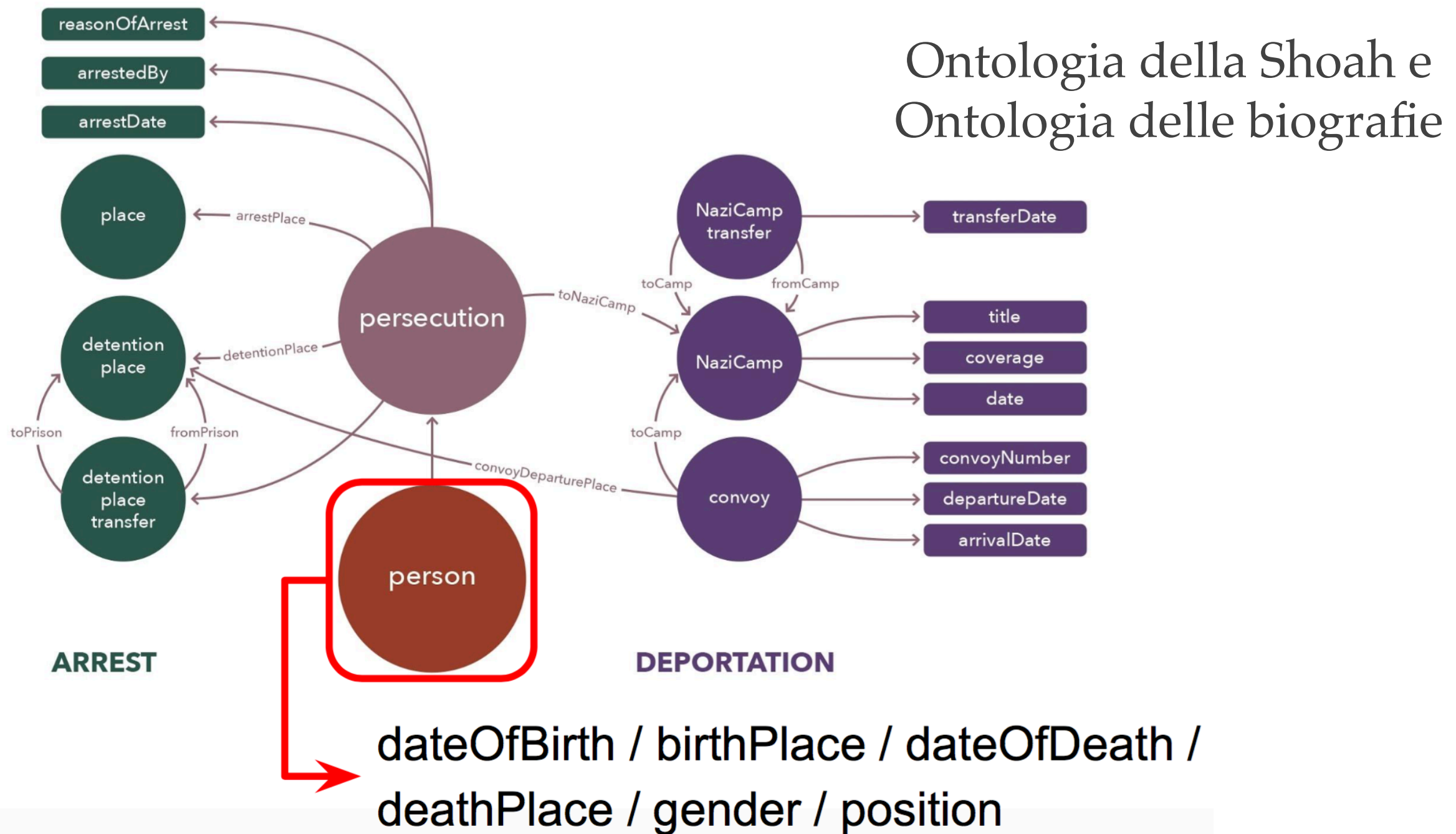
persecution of Levi, Primo 31/7/1919 Torino

<http://dati.cdec.it/lod/shoah/persecution/5002>

ENTITÀ DI TIPO: Persecution

rdfs:label	persecution of Levi, Primo 31/7/1919 Torino
shoah:convoyNumber	8
shoah:departureDate	22/02/1944
shoah:arrivalDate	26/02/1944
shoah:campSerialNumber	174517
shoah:persecutionCategory	ebreo
dc:abstract	EN IT Primo Levi, figlio di Cesare Levi e Ester Luzzati è nato in Italia a Torino il 31 luglio 1919. Arrestato a Brusson (Aosta). Deportato nel campo di sterminio di Auschwitz. È sopravvissuto alla Shoah.
shoah:arrestedBy	italiani
shoah:arrestDate	13/12/1943
rdf:type	shoah:Persecution ↳ Persecution

Osservazione dei dati



Schema preso da “Shoah vocabulary Specification” <http://dati.cdec.it>

Estrazione dei dati

Virtuoso SPARQL Query Editor

1	ID	name	gender	date_of_birth	place_of_birth	date_of_death	place_of_death	massacreDate
554	http://dati.cdec.it/lod/shoah/person/5894	Ninos, Luisa	F	19050830	Smirne			
555	http://dati.cdec.it/lod/shoah/person/5894	Ninos, Luisa	F	19050830	Smirne			
556	http://dati.cdec.it/lod/shoah/person/5002	Levi, Primo	M	19190731	Torino	19870411	Torino	
557	http://dati.cdec.it/lod/shoah/person/2363	Di Gioacchino, Anna	F	19110120	Ancona	19480413	Gerusalemme	
558	http://dati.cdec.it/lod/shoah/person/2363	Di Gioacchino, Anna	F	19110120	Ancona	19480413	Gerusalemme	
559	http://dati.cdec.it/lod/shoah/person/2363	Di Gioacchino, Anna	F	19110120	Ancona	19480413	Gerusalemme	
560	http://dati.cdec.it/lod/shoah/person/1818	Coné, Giuseppe	M	19100207	Rodi			
561	http://dati.cdec.it/lod/shoah/person/1818	Coné, Giuseppe	M	19100207	Rodi			
562	http://dati.cdec.it/lod/shoah/person/1818	Coné, Giuseppe	M	19100207	Rodi			
563	http://dati.cdec.it/lod/shoah/person/658	Arbisse, Raimondo	M	19190219	Minsk			
564	http://dati.cdec.it/lod/shoah/person/4980	Levi, Felicia	F	19230313	Rodi			
565	http://dati.cdec.it/lod/shoah/person/4980	Levi, Felicia	F	19230313	Rodi			
566	http://dati.cdec.it/lod/shoah/person/4980	Levi, Felicia	F	19230313	Rodi			
567	http://dati.cdec.it/lod/shoah/person/5014	Levi, Vittoria	F	19200110	Rodi			
568	http://dati.cdec.it/lod/shoah/person/5014	Levi, Vittoria	F	19200110	Rodi			
569	http://dati.cdec.it/lod/shoah/person/3281	Goldstein, Bronia Beatrice	F	19080215	Il Cairo			
570	http://dati.cdec.it/lod/shoah/person/5895	Nissim, Luciana	F	19191020	Torino	19981201	Milano	
571	http://dati.cdec.it/lod/shoah/person/5895	Nissim, Luciana	F	19191020	Torino	19981201	Milano	
572	http://dati.cdec.it/lod/shoah/person/1820	Cordoval, Eliakim	M	19190115	Rodi			
573	http://dati.cdec.it/lod/shoah/person/1820	Cordoval, Eliakim	M	19190115	Rodi			
574	http://dati.cdec.it/lod/shoah/person/1820	Cordoval, Eliakim	M	19190115	Rodi			
575	http://dati.cdec.it/lod/shoah/person/213	Alhadeff, Samuele	M	19070407	Rodi	19450410		
576	http://dati.cdec.it/lod/shoah/person/850	Berger, Giuseppe	M	19210215	Zaluzice	19450302		
577	http://dati.cdec.it/lod/shoah/person/714	Bassani, Bruno	M	19121230	Bologna	19450126		
578	http://dati.cdec.it/lod/shoah/person/714	Bassani, Bruno	M	19121230	Bologna	19450126		

(The result can only be sent back to browser, not saved on the server, see [details](#))

Run Query Reset

Da Query Sparql a CSV

Estrazione dei dati

1. ID	14. shoahSurvivor
2. name	15. arrestDate
3. gender	16. arrestPlace
4. date_of_birth	17. departureDate
5. place_of_birth	18. convoyDeparturePlace
6. date_of_death	19. convoyNumber
7. place_of_death	20. convoyArrivalDate
8. massacreDate	21. arrivalDate
9. massacrePlace	22. toNaziCamp
10. deathDescriptionIntegration	23. naziCampTransfertDate
11. position	24. toNaziCampTranfert
12. abstract	25. liberationReturnDate
13. deathDescription	26. liberationReturnPlace

Informazioni relative a 9.042 persone

Integrazione e correzione

date_of_birth	arrivalDate	naziCampTransfertDate
19250315	?	00/12/1944
1900	19440816	1945.04
19220501	19440816	18/01/1945
0	?	1944.02.26
18830807	16/06/1944	?
19270806	?	02/02/1945
19448	7 marzo 1945	1944.10. 27
1917	00/00/1944	?*



14 formati per esprimere una data,
tutti convertiti in
AAAA-MM-GG

Data mancante =
No spostamento

Integrazione e correzione

Professioni uniformate e semplificate, da 168 categorie a 27 utilizzando la classificazione di **Wikipedia**

Category:Occupations by type

From Wikipedia, the free encyclopedia

Subcategories

This category has the following 36 subcategories, out of 36 total.

*

- ▶ [Indian castes by profession](#) (18 C)

A

- ▶ [Agricultural occupations](#) (5 C, 3 P)
- ▶ [Arts occupations](#) (11 C, 34 P)

C

- ▶ [Ceremonial occupations](#) (10 P)
- ▶ [Cleaning and maintenance occupations](#) (2 C, 30 P)
- ▶ [Computer occupations](#) (4 C, 45 P)
- ▶ [Construction and extraction occupations](#) (4 C, 16 P)
- ▶ [Consulting occupations](#) (1 C, 32 P)

D

- ▶ [Design occupations](#) (5 C, 11 P)

E

- ▶ [Economic occupations](#) (3 C, 5 P)
- ▶ [Education and training occupations](#) (6 C, 77 P)
- ▶ [Engineering occupations](#) (6 C, 27 P)
- ▶ [Entertainment occupations](#) (17 C, 70 P)

F

- ▶ [Fashion occupations](#) (6 C, 27 P)

G

- ▶ [Gendered occupations](#) (10 C, 62 P)
- ▶ [Government occupations](#) (10 C, 112 P)

H

- ▶ [Healthcare occupations](#) (26 C, 157 P)
- ▶ [Hospitality occupations](#) (2 C, 16 P)
- ▶ [Humanities occupations](#) (2 C, 10 P)

I

- ▶ [Illegal occupations](#) (11 C, 34 P)
- ▶ [Industrial occupations](#) (8 C, 19 P)

J

- ▶ [Journalism occupations](#) (3 C, 27 P)

L

- ▶ [Legal professions](#) (13 C, 183 P)

M

- ▶ [Maritime occupations](#)
- ▶ [Marketing occupations](#)
- ▶ [Media occupations](#)
- ▶ [Military specialism](#)

O

- ▶ [Ombudsmen by type](#)

P

- ▶ [Political occupations](#)
- ▶ [Product testers](#) (2)
- ▶ [Production occupations](#)

R

- ▶ [Religious occupations](#)

S

- ▶ [Science occupations](#)
- ▶ [Service occupations](#)
- ▶ [Sports occupations](#)

T

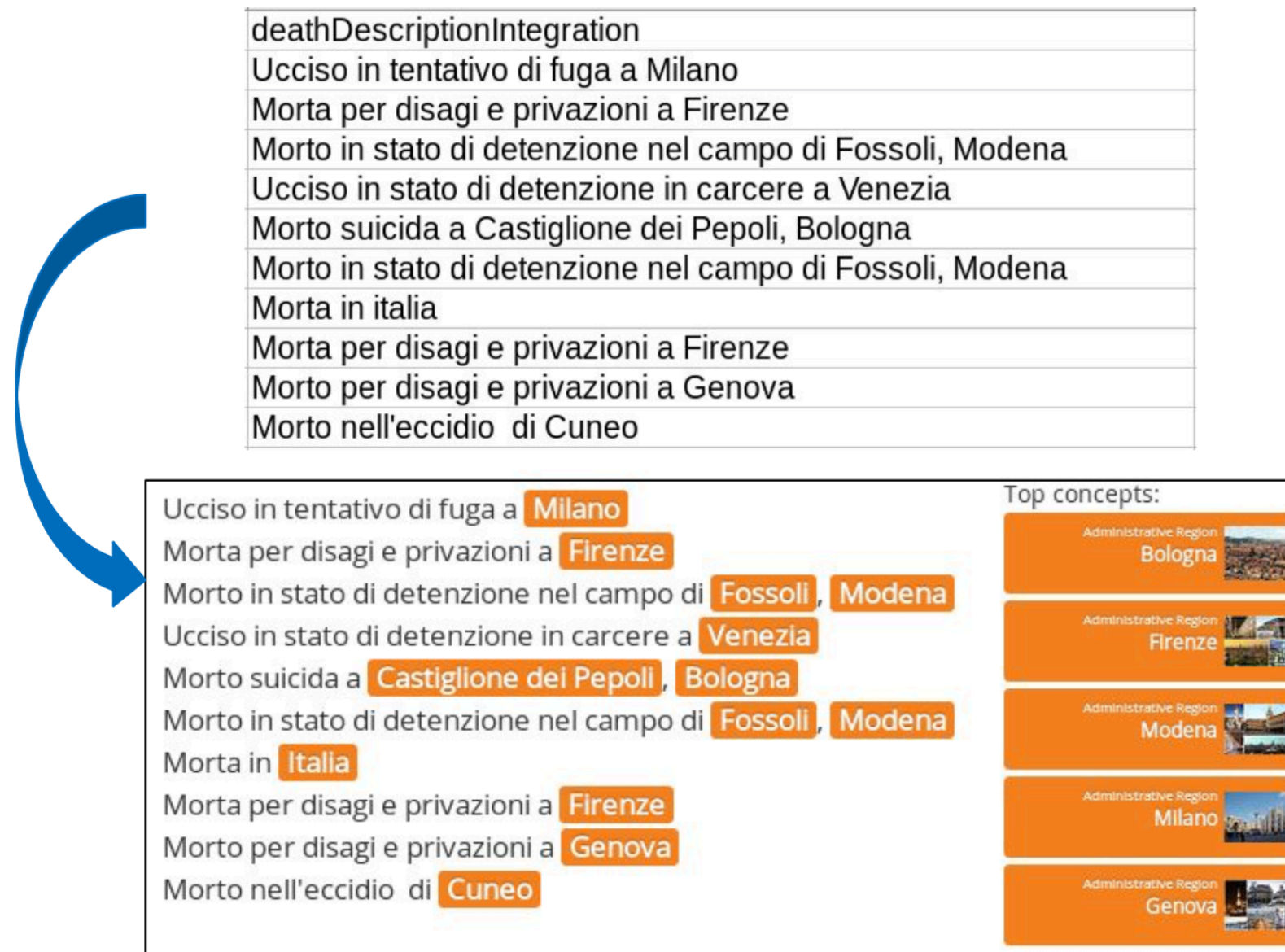
- ▶ [Transport occupations](#)



WIKIPEDIA
The Free Encyclopedia

Integrazione e correzione

Integrazione di luoghi mancanti: 312 località non specificate nel campo *PlaceOfDeath* recuperate in modo automatico processando il campo *DeathDescriptionIntegration*



Integrazione e correzione

Integrazione di luoghi mancanti: *CountryOfOrigin* mancante recuperata automaticamente dal campo *Abstract*

abstract
Elena Recanati, daughter of Luigi Recanati and Luigia Simon. She was born in Italy , in Turin, on March 12 1922. She was married with Guido Foa. She was arrested in Canischio (Turin). She was deported to the extermination camp of Auschwitz. She survived the Shoah.
Regina Assael, daughter of Menahem Assael and Emilia Politi. She was born in Turkey , in Izmir, on June 14 1872. She was the widow of Isaac Ventura. She was arrested in Milan (Milan). She was deported to Bergen Belsen. She survived the Shoah.
Erich Hirschl, son of Hinko Hirschl and Giulia Hirsch. He was born in Yugoslavia in Zagreb, on March 21 1921. He was married with Zora Adler. He was arrested in Milan (Milan). He was deported to the extermination camp of Auschwitz. He did not survive the Shoah.

Integrazione e correzione

Georeferenziazione delle località usando il tool **Nominatim**: 1.493 località, il 15% sono state corrette o integrate manualmente

Es. Gross-Meseritz -> Gross Meseritsch
Choumbla -> Bulgaria

place_of_birth
Rodi
Sarajevo
Roma
Torino
Fiume
Smirne
Trieste
Homs

 Nominatim

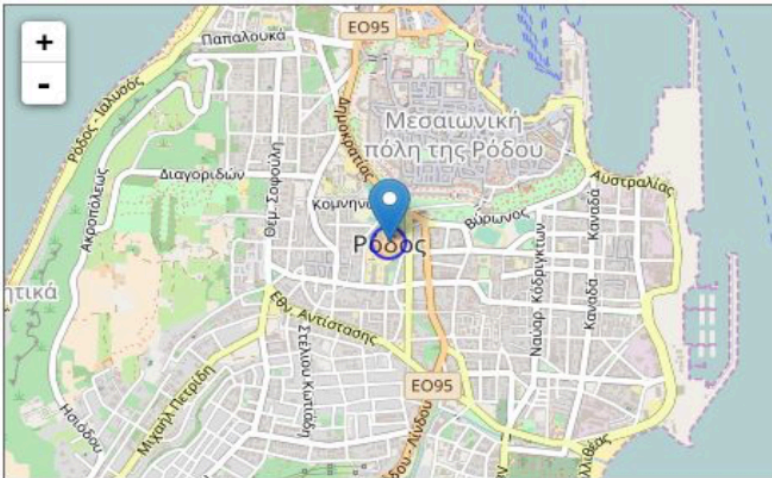
Rhodes

Name	Rodos (int_name) Ρόδος (name) Rhodos (name:cs) Rhodos (name:de) Ρόδος (name:el) Rhodes (name:en) Rodas (name:es) Rodos (name:fi) Rhodes (name:fr) Rodosz (name:hu) Rodi (name:it) Rodos (name:pl) Rodos (name:ro) Ρόδος (name:ru) Rodos (name:sk) Ρόδος (name:sr) Rodos (name:sr-Latn) Rhodos (name:sv) Rodi (old_name)
Type	place:city
Last Updated	2017-06-04 21:21
Admin Level	15
Rank	City
Importance	0.35 (estimated)
Coverage	Point
Centre Point	36.4379874,28.2233083

Data last updated:
2017/06/21 12:04 GMT

About & Help ▾





Integrazione e correzione

Conversione dei dati in formato json

```
{
  "name": "Primo_Levi-5002",
  "nation": "Italy",
  "continent": "EUROPE",
  "sex": "M",
  "work": "Mixed",
  "URI": "http://dati.cdec.it/lod/shoah/person/5002",
  "type": "Survivor of Shoah",
  "movements": [
    {
      "date": 19190731,
      "place": "Torino",
      "latitude": 45.0709201,
      "longitude": 7.685972,
      "predicate_id": "Unknown",
      "predicate": "Unknown",
      "resource": "CDEC Data",
      "resource_frame": "Birth",
      "place_frame": "Unknown",
      "snippet": "Born in Torino in/on July 31, 1919"
    },
    {
      "date": 19431213,
      "place": "Brusson",
      "latitude": 45.759342,
      "longitude": 7.7304474,
      "predicate_id": "Unknown",
      "predicate": "Unknown",
      "resource": "CDEC Data",
      "resource_frame": "Arrest",
      "place_frame": "Unknown",
      "snippet": "Arrested in Brusson in/on December 13, 1943"
    }
  ]
}
```

Visualizzazione

Interfaccia interattiva:

- Ricerca per nome o luogo
- Linea del tempo
- Filtri: *country of origin / occupation / death_description / gender*
- Link alla libreria digitale del CDEC

Visualizzazione

[Demo LODNavigator]

LOD Navigator

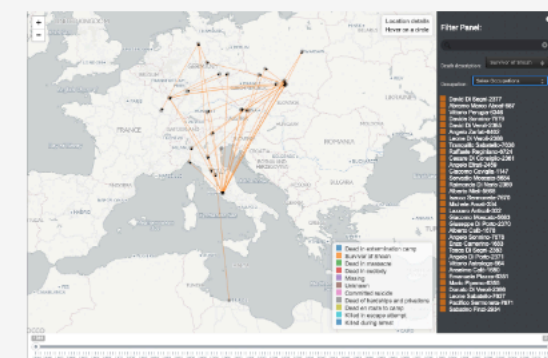
The LOD NAVIGATOR takes in input the [data](#) made available by the Contemporary Jewish Documentation Center (CDEC) in Milan, in collaboration with [regesta.exe](#), and published in Linked Open Data format to show the movements of the Italian victims of Shoah. To this end, we used the SPARQL endpoint to collect biographical data together with information about the persecution and deportation of each victim. The places of birth, arrest, detention, deportation to a Nazi camp, transfer, and return after liberation (if available) were georeferenced and associated to the corresponding date. All the movements are then displayed in an interactive interface.

The tool is distributed in different releases complying with different operating systems and with a CC-BY-NC-ND 4.0 license.

Download: http://dhlab.fbk.eu/lod_navigator/

Reference:

Rachele Sprugnoli, Giovanni Moretti, Sara Tonelli. *LOD Navigator: Tracing Movements of Italian Shoah Victims*. To appear in Proceedings of EHRI workshop Data Sharing, Holocaust Documentation, Digital Humanities.



Documents:



Analisi quantitativa

Cosa viene visualizzato:

- 8,936 persone, 18,813 movimenti tra 37,626 località

death_description	QUANTITY
Dead in extermination camp	7,333
Survivor of Shoah	1,037
Dead in massacre	202
Dead in custody	70
Unknown	28
Dead en route to camp	15
Committed suicide	10
Dead of hardships and privations	6
Killed in escape attempt	6
Missing	6
Killed during arrest	1
TOTAL	8,712

Statistiche in base a death_description

Analisi quantitativa

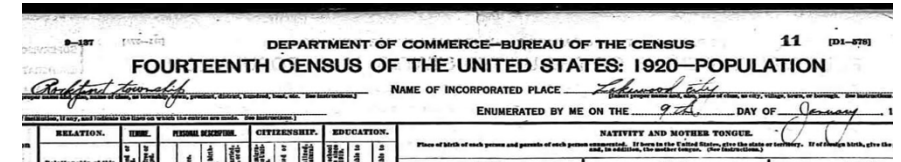
Cosa viene visualizzato:

- 8.936 persone, 18.813 movimenti tra 37.626 località

EVENTS	#	TOP 5 PLACES
Birth	8,340	Roma (1,526) - Rhodes (1,524) - Trieste (466) - Vienna (205) - Venezia (200)
Arrest	7,960	Rhodes (1,758) - Roma (1,712) - Trieste (509) - Borgo San Dalmazzo (329) - Milano (241)
Detention	7,873	Fossoli (2,699) - Rhodes (1,712) - Roma (1,029) - Milano (872) - Trieste (661)
To Nazi Camp	7,893	Auschwitz (7,338) - Bergen Belsen (405) - Ravensbrueck (89) - Flossenbug (27) - Buchenwald (25)
Nazi Camp Transfer	821	Vittel (107) - Monowitz (86) - Biberach (75) - Buchenwald (59) - Mauthausen (56)
Death	4,332	Auschwitz (3,743) - Roma (81) - Monowitz (57) - Lago Maggiore (50) - Flossenbug (38)
Return after Liberation	136	Roma (86) - Milano (32) - Torino (7) - Tripoli (4) - Livorno (2)
TOTAL	37,355	

“Spatial Humanities” – dalla carta al computer

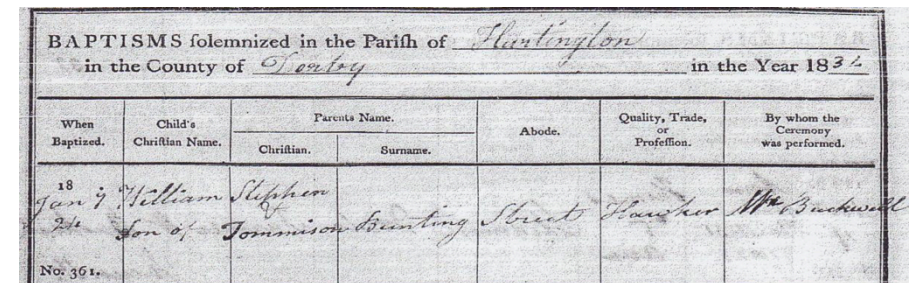
Approccio tradizionale



Estrazione di informazione spaziale tramite:

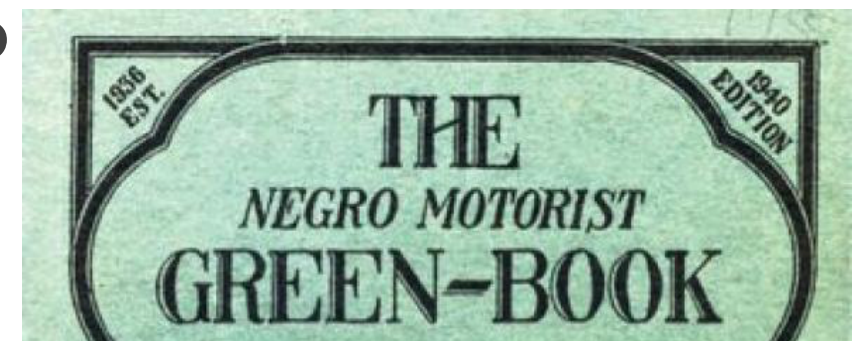
- Fonti primarie strutturate

Es. Registro cartaceo delle vittime italiane della Shoah



- Libri, diari, resoconti, racconti di viaggio

Es. Articoli sulle eruzioni di vulcani

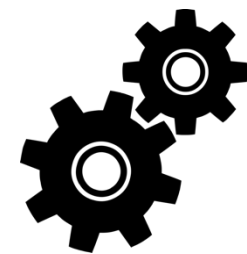


“Spatial Humanities” – dalla carta al computer

Approccio “digitale”

Estrazione di informazione spaziale tramite:

- Integrazione e armonizzazione di LOD, database
- Estrazione automatica di informazioni usando sistemi per **l'analisi del testo** (NLP = Natural Language Processing)



Moduli per l'Analisi del Testo

Il lupo perde il pelo ma non il vizio

Moduli per l'Analisi del Testo

Il lupo perde il pelo ma non il vizio

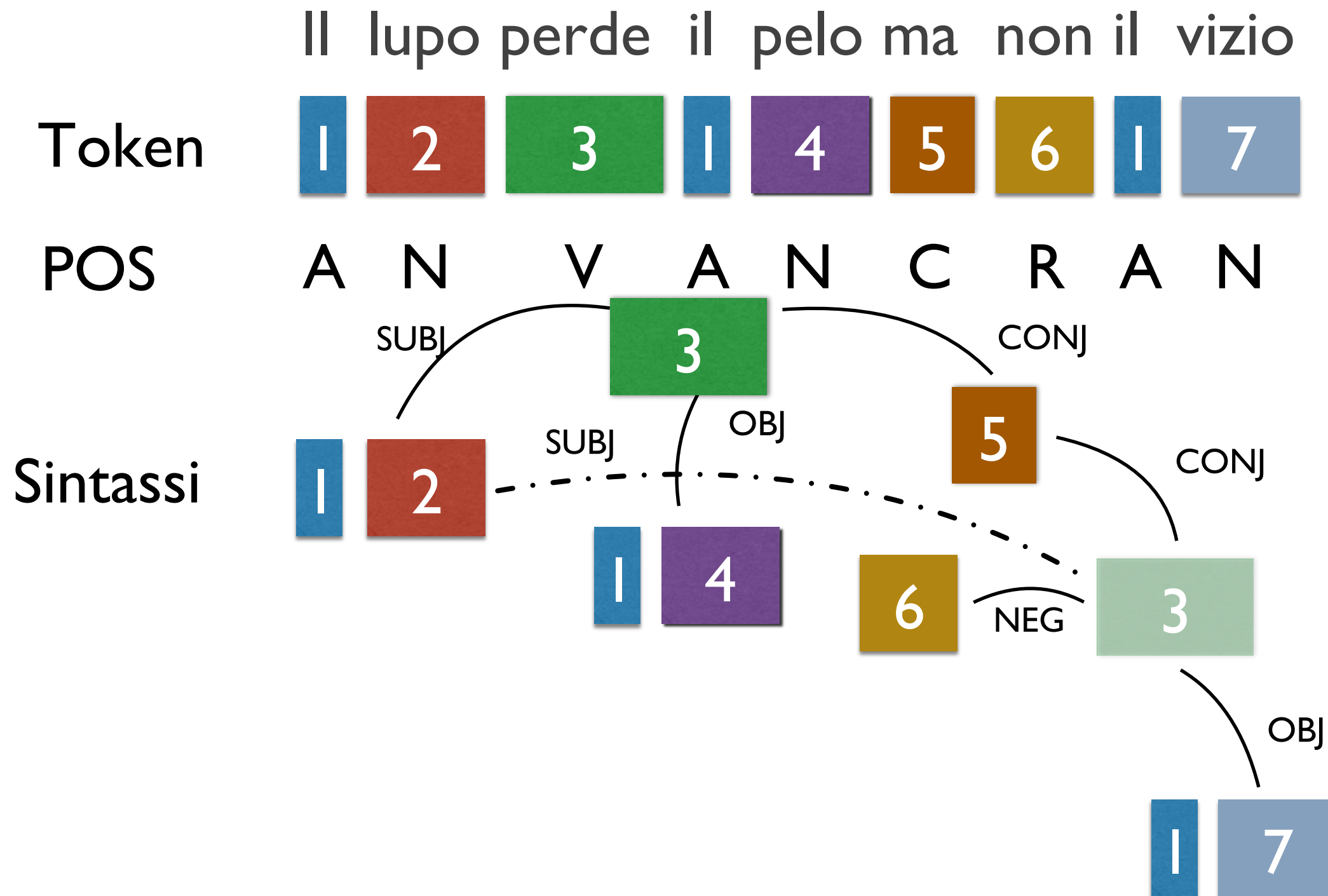
Token



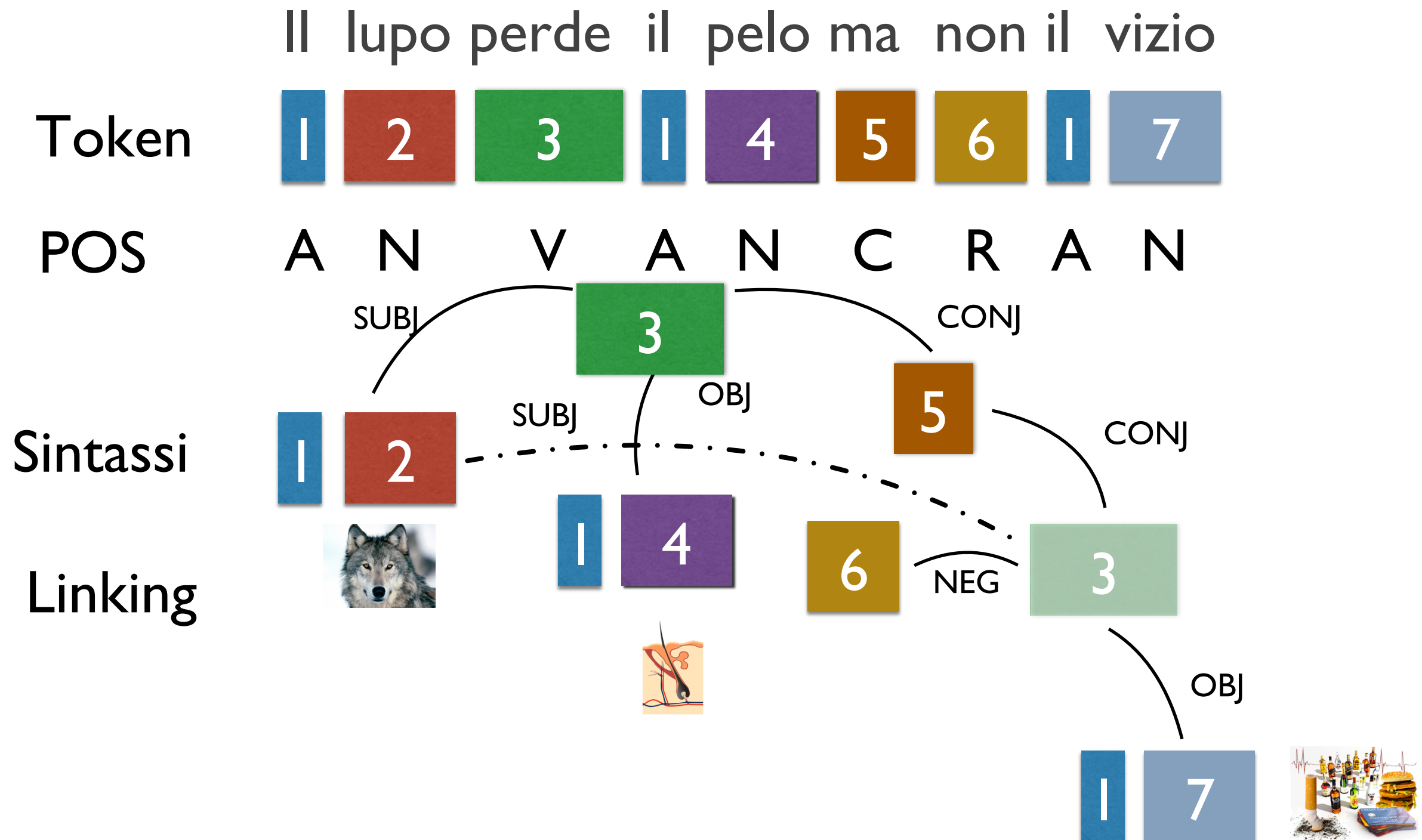
Moduli per l'Analisi del Testo

	Il lupo perde il pelo ma non il vizio								
Token	1	2	3	1	4	5	6	1	7
POS	A	N	V	A	N	C	R	A	N

Moduli per l'Analisi del Testo



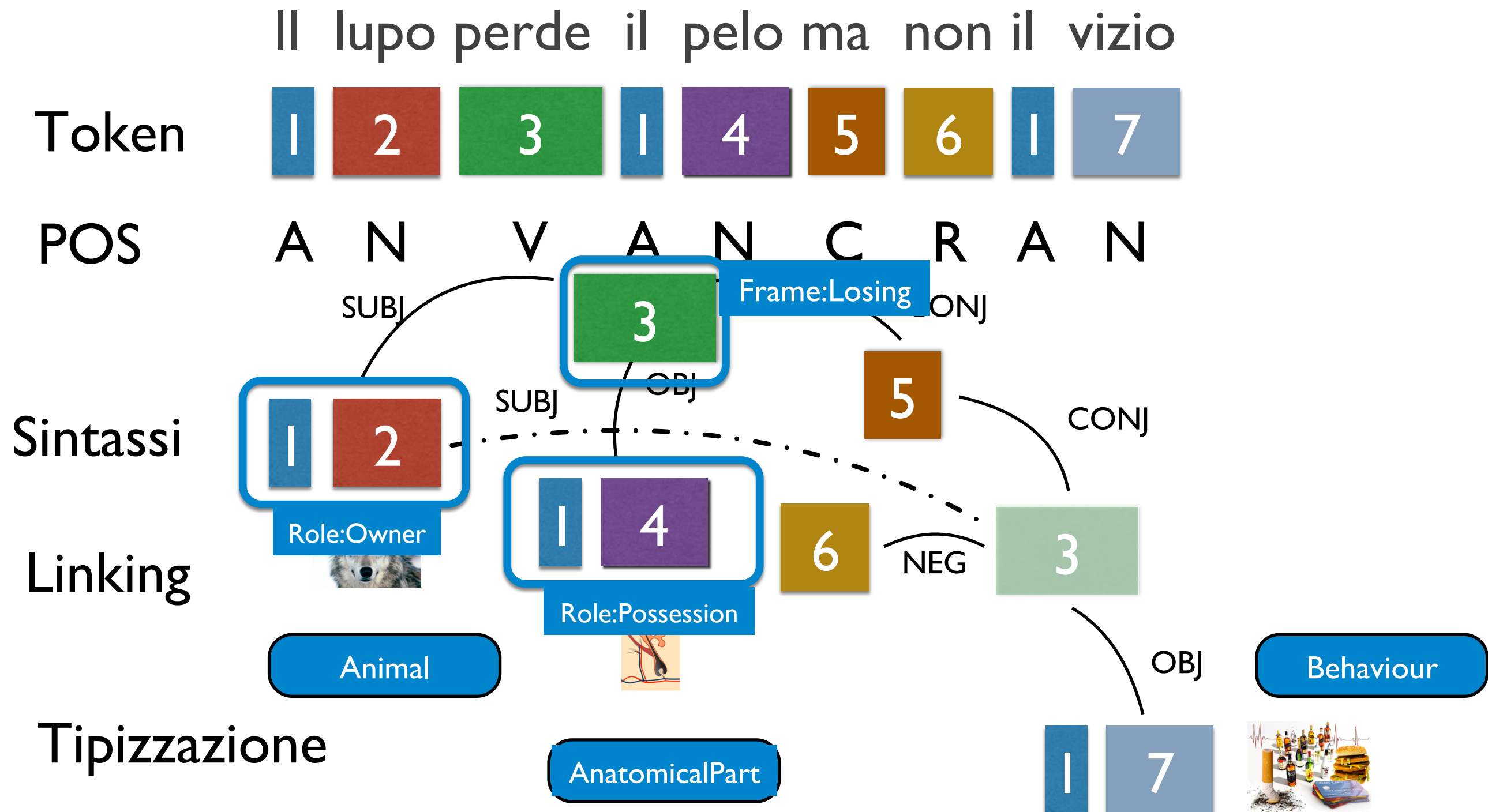
Moduli per l'Analisi del Testo



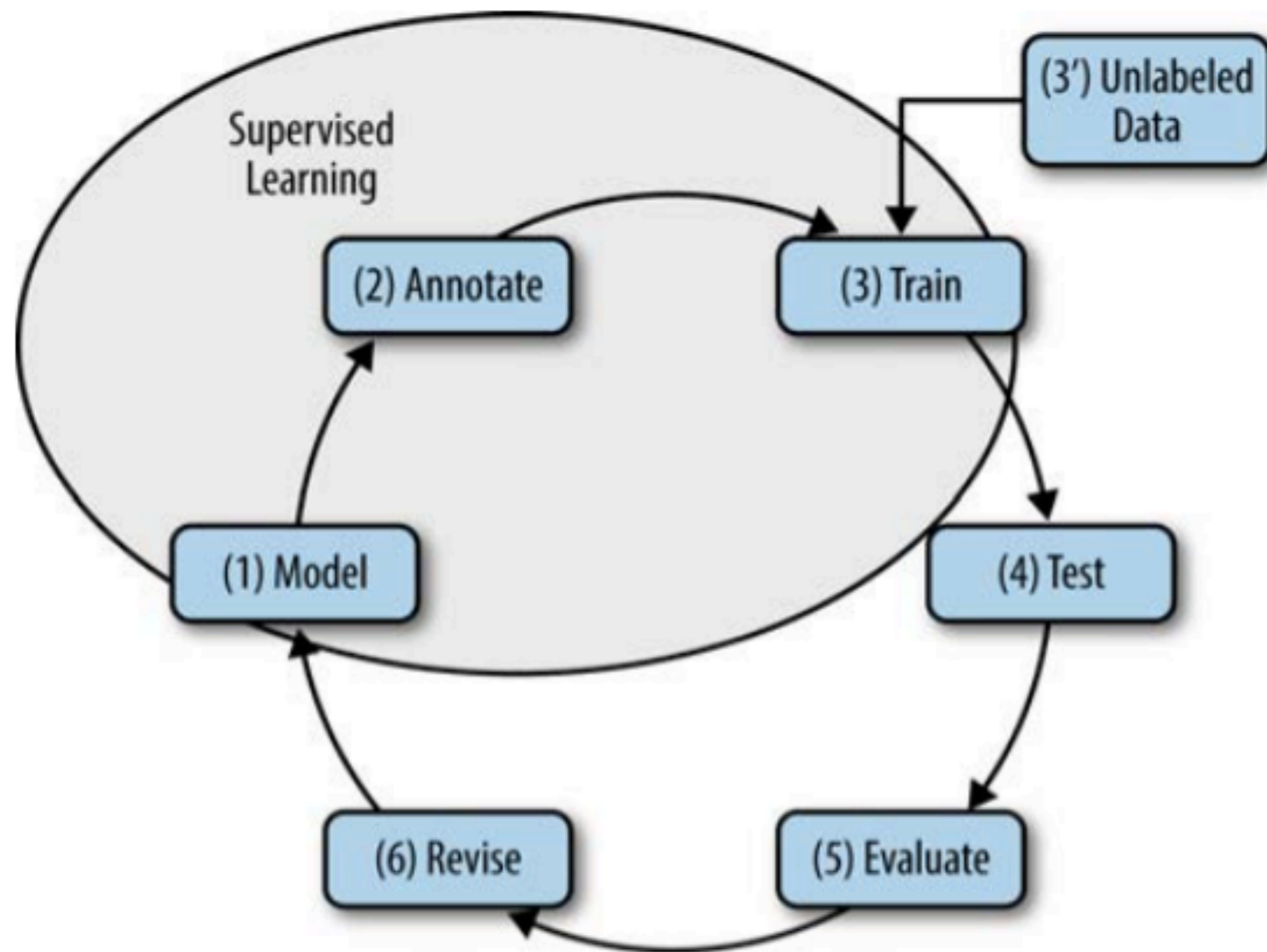
Moduli per l'Analisi del Testo



Moduli per l'Analisi del Testo



Come si crea un modulo per l'analisi del testo?



Dal modello ai dati annotati ai sistemi automatici

(Pustejovsky and Stubbs, 2012)

Come funziona un modulo per l'analisi del testo?

[Demo TINT

<http://tint.fbk.eu>]

Dal modello ai dati annotati ai sistemi automatici

(Pustejovsky and Stubbs, 2012)

Progetto ALCIDE

- Collezione completa degli scritti pubblici di Alcide De Gasperi (1881 – 1954)
- Circa 3.000 documenti contenenti 3.000.000 di token (~70.000 lemma)
- Comprende trascrizioni di discorsi, articoli di giornale, atti parlamentari, telegrammi, ecc.
- Versione cartacea raccolta in 7 tomi pubblicata da ISIG



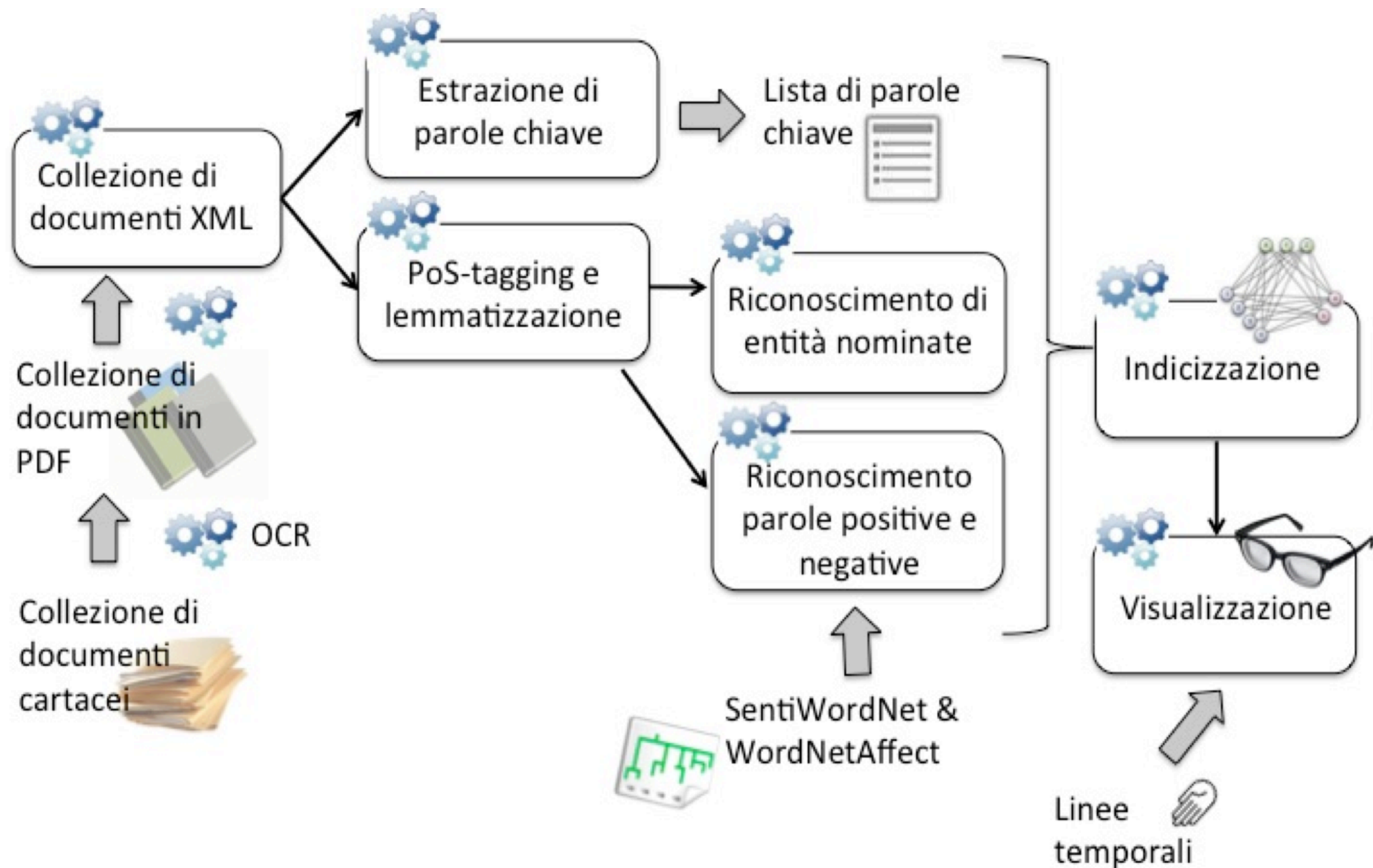
Progetto ALCIDE: Requisiti

- Possibilità di esplorare il *corpus* selezionando il periodo temporale di riferimento
- Il *periodo temporale* non deve essere predefinito
- Al termine di ogni ricerca deve essere possibile visualizzare il contenuto dei documenti trovati: evitare visualizzazioni statiche
- Possibilità di *annotare* e *esportare* i risultati della ricerca
- Creare un'infrastruttura non legata a un particolare corpus o periodo, renderla più flessibile possibile

Progetto ALCIDE: Requisiti

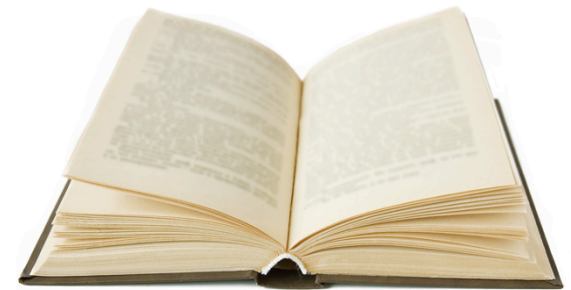
- Possibilità di aggiungere facilmente nuovi corpora, utilizzo di formati standard (XML) per consentire l'uso a utenti senza competenze tecniche
- Efficienza per consentire ricerche in tempo reale, velocità
- Eliminare lavoro manuale, architettura modulare automatizzata
- Consultabile tramite browser, evitare applicazioni da installare
- Evitare visualizzazioni troppo complesse, HTML5

Sistema di NLP



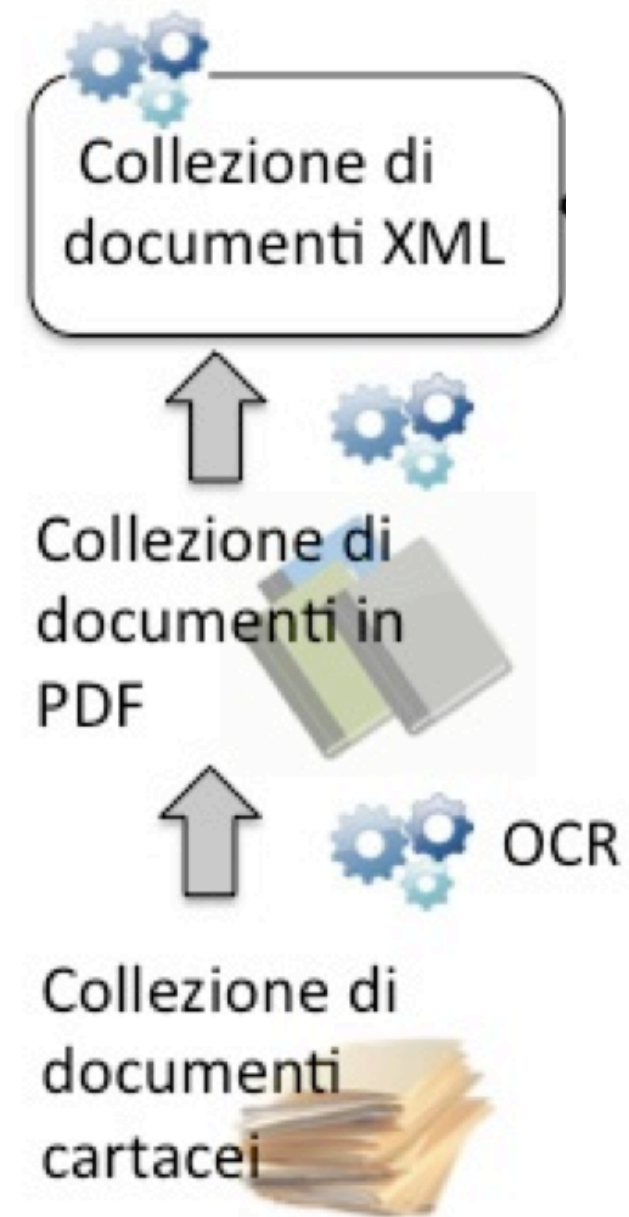
Sistema di NLP

Dobbiamo fare leggere i documenti al computer, quindi devono essere **scritti in modo comprensibile** per il computer



```
<file id="1">
  <head>
    <url>I.pdf_doc_number_1</url>
    <date>1901-01-01</date>
    <title>Title of the document</title>
    <description>Description of the Document</description>
    <location>Geographic location</location>
    <abstract>Content of the abstract</abstract>
  </head>
  <content>Text of the document in Italian</content>
  <original_text>Text of the document not in Italian - if present</original_text>
</file>
```

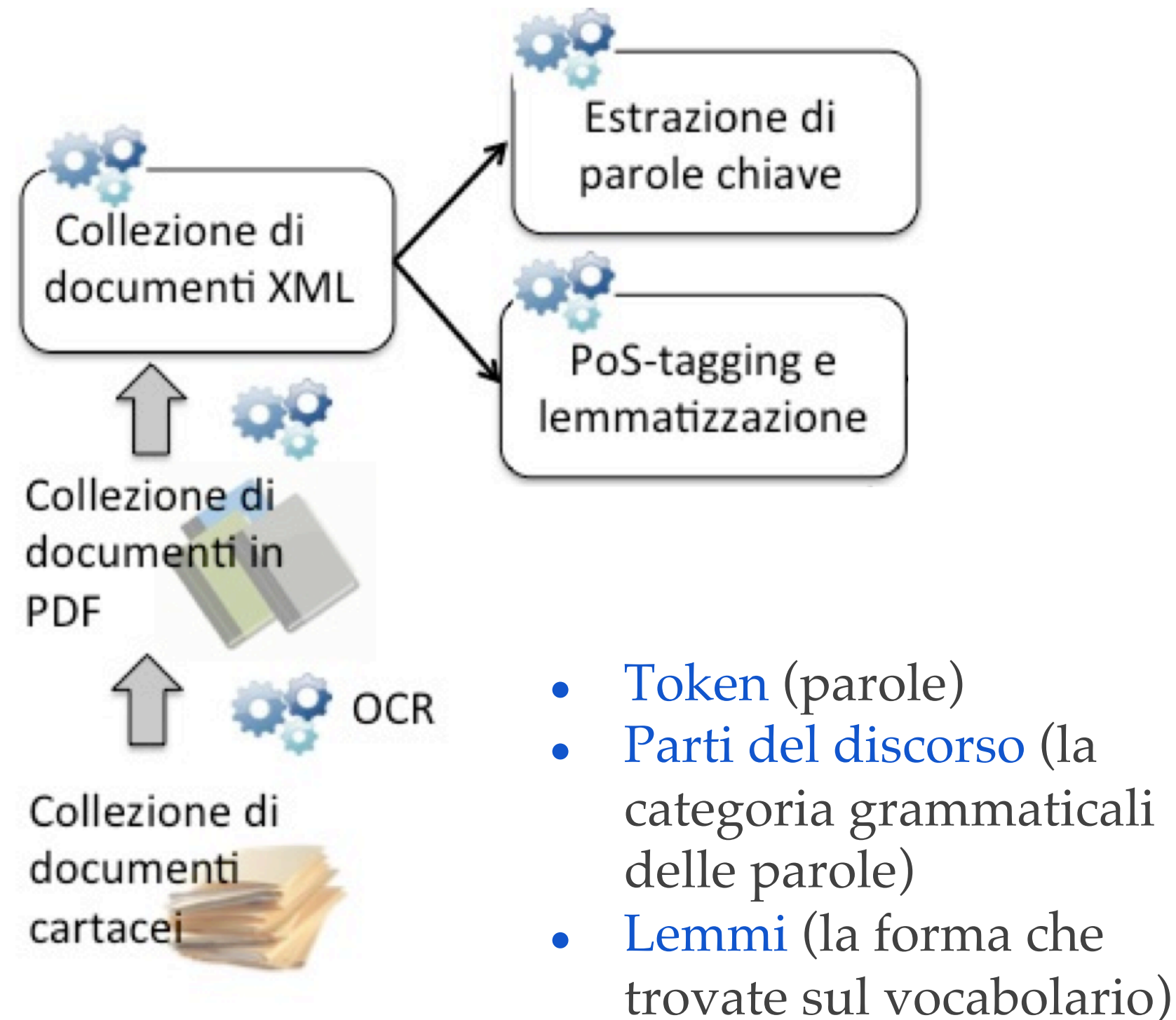
Sistema di NLP



De Gasperi nacque a Pieve Tesino in una povera famiglia tirolese : infatti i suoi genitori dovettero chiedere un sussidio allo Stato austriaco per farlo studiare . Era il primo dei quattro figli di Maria Morandini , nata a Predazzo , e Amedeo De Gasperi , nato a Sardagna.

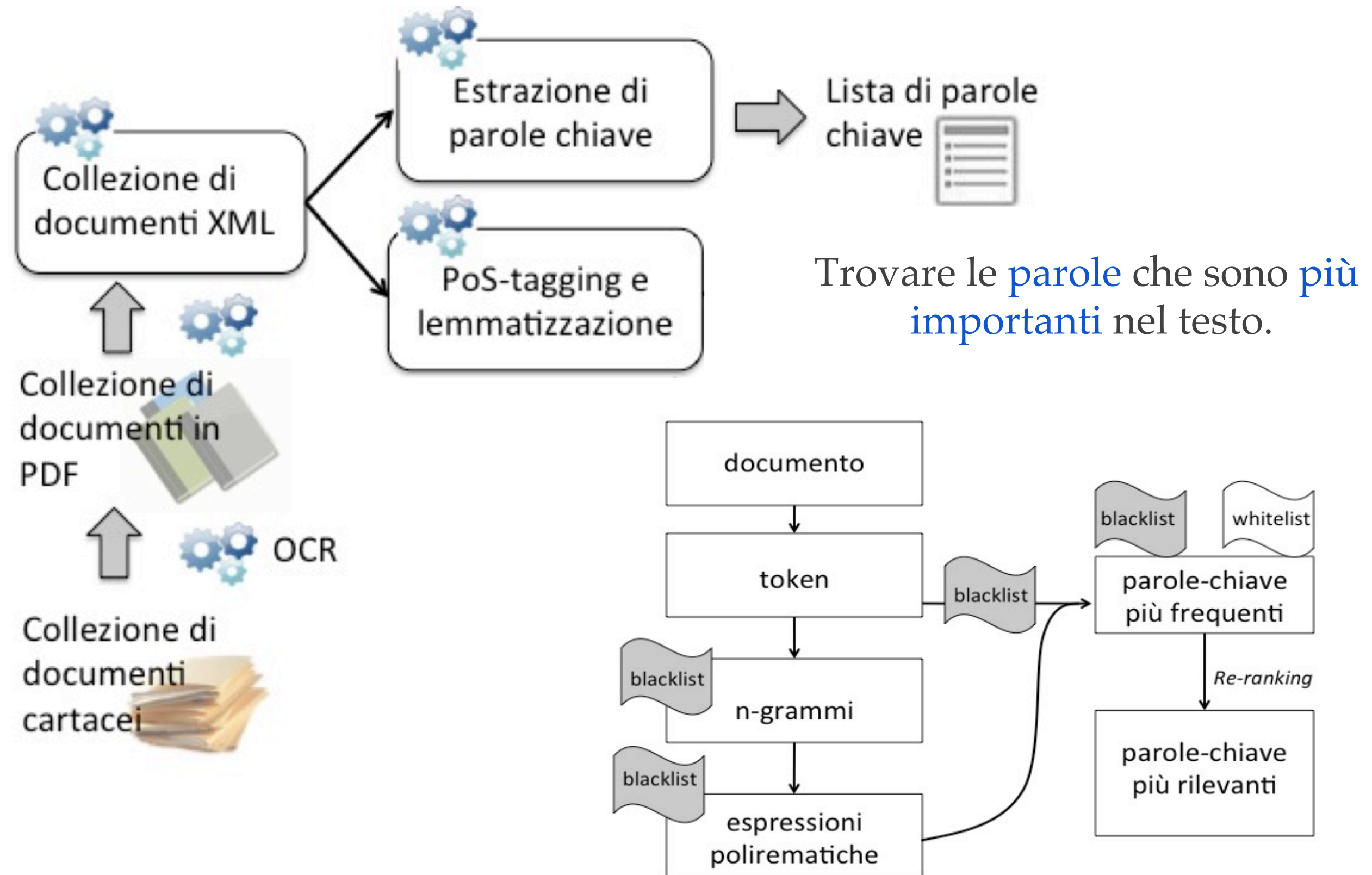
Un computer **capisce** questa frase?

Sistema di NLP

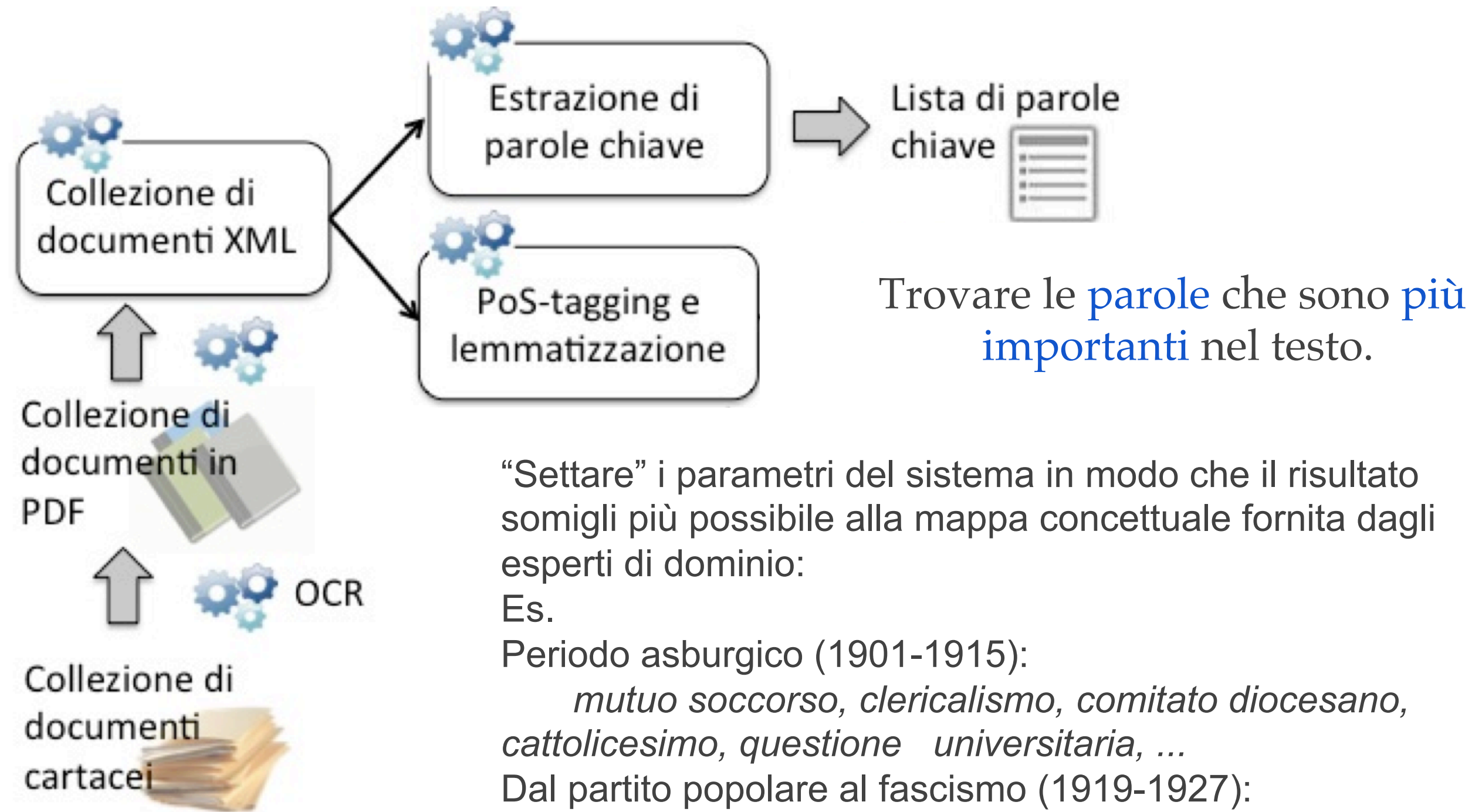


token	pos	lemma
De	SPN	__NULL__
Gasperi	SPN	__NULL__
nacque	VI	nascere
a	E	a
Pieve	SPN	__NULL__
Tesino	SPN	__NULL__
in	E	in
una	RS	indet
povera	AS	povero
famiglia	SS	famiglia
tirolese	AS	tirolese

Sistema di NLP



Sistema di NLP



“Settare” i parametri del sistema in modo che il risultato somigli più possibile alla mappa concettuale fornita dagli esperti di dominio:

Es.

Periodo asburgico (1901-1915):

mutuo soccorso, clericalismo, comitato diocesano, cattolicesimo, questione universitaria, ...

Dal partito popolare al fascismo (1919-1927):

profughi, socialismo, ricostruzione, guerra, pace, Versailles, regionalismo, Trentino, ...

Sistema di NLP



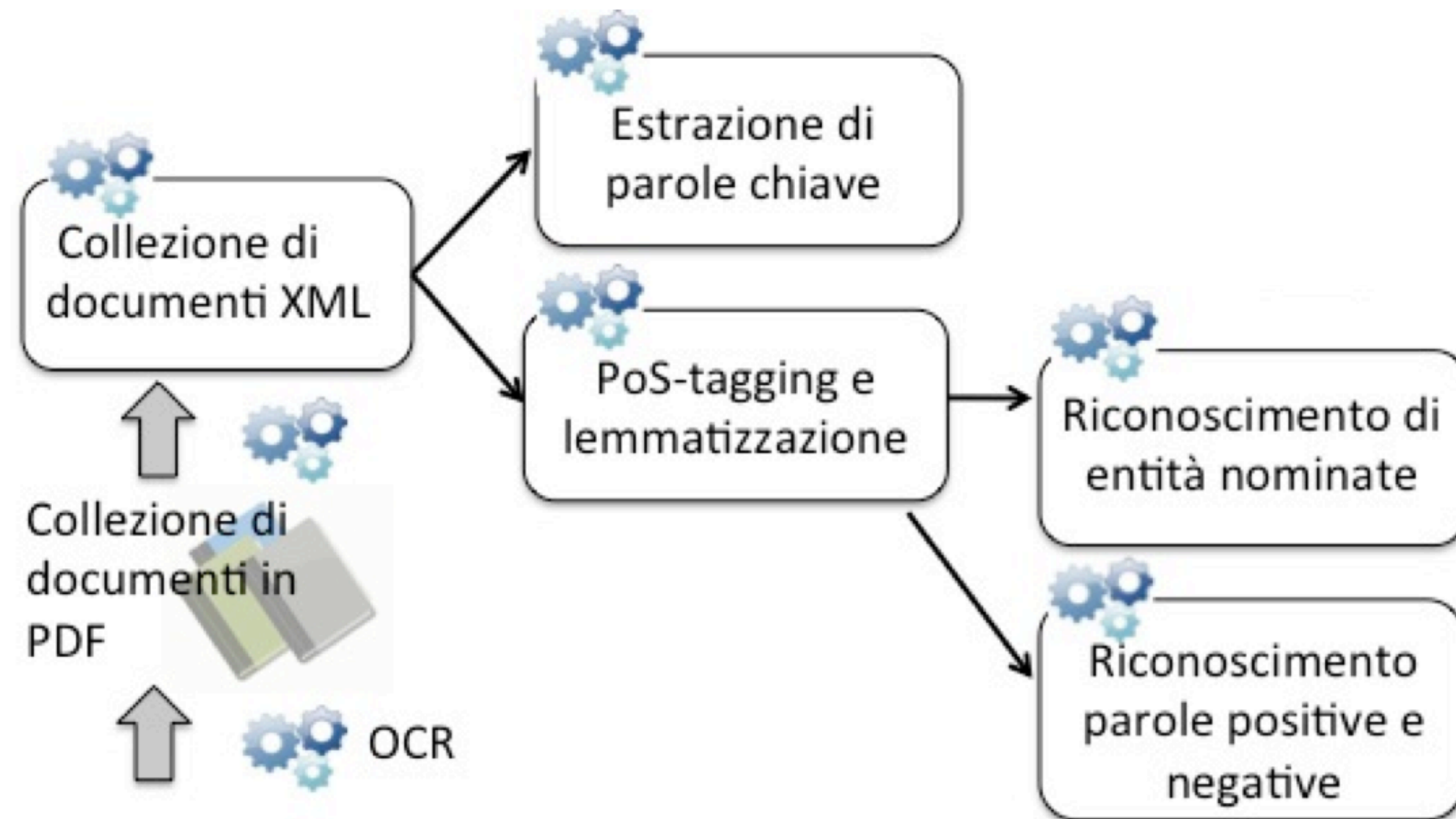
Entità nominate: persone o cose che hanno un nome proprio.

☒ Person ☐ Organization ☒ Location

De Gasperi nacque a Pieve Tesino in una povera famiglia tirolese : infatti i suoi genitori dovettero chiedere un sussidio allo Stato austriaco per farlo studiare .
Era il primo dei quattro figli di Maria Morandini , nata a Predazzo , e Amedeo Degasperi , nato a Sardinia .

Dopo di lui nacquero Mario , che fu sacerdote , Marcella e Augusto .

Sistema di NLP



"De Gasperi nacque a Pieve Tesino in una povera famiglia tirolese"

Sistema di NLP



Sistema di NLP

[DEMO ALCIDE]

http://celct.fbk.eu:8080/Alcide_Demo/login

ALCIDE: Valutazione tramite focus group

- Alcune funzionalità (es. *Timeline* o link delle entità nominate a *Wikipedia*) sono state rimosse perchè considerate più adatte per scopi didattici
- L'estrazione dei concetti-chiave è un processo dai criteri “soggettivi”, utile ma di difficile interpretazione
- La polarità a livello di documento è poco utile, è stata rimossa
- Sarebbe utile poter avere un'analisi di come cambia il significato di una parola nel tempo (per es. guardando le co-occorrenze)

ALCIDE: Valutazione automatica

	News	De Gasperi Corpus
	Accuracy	Accuracy
PoS	0.96	0.95

	News			De Gasperi corpus		
	P	R	F1	P	R	F1
PER	0.92	0.93	0.92	0.70	0.82	0.76
ORG	0.69	0.60	0.64	0.23	0.39	0.29
LOC	0.78	0.69	0.73	0.50	0.50	0.50
GPE	0.85	0.86	0.85	0.82	0.90	0.86
TOTAL	0.83	0.80	0.82	0.62	0.76	0.69

Valutazione automatica

	News			Historical corpus		
	P	R	F1	P	R	F1
Italian	0.88	0.85	0.87	0.89	0.78	0.83
Spanish	0.92	0.86	0.89	0.27	0.72	0.39

Annotazione di “eventi”

Cosa abbiamo imparato sulla interdisciplinarietà

- **Non basta lavorare insieme e avere background diversi per parlare di interdisciplinarietà**
 - Avere obiettivi chiari
 - Affrontare diversi cicli di interazione
 - Adottare un linguaggio comune
 - Non essere impazienti
 - Essere aperti alle novità e disponibili a modificare il proprio metodo di lavoro

Grazie per l'attenzione

Sara Tonelli

satonelli@fbk.eu