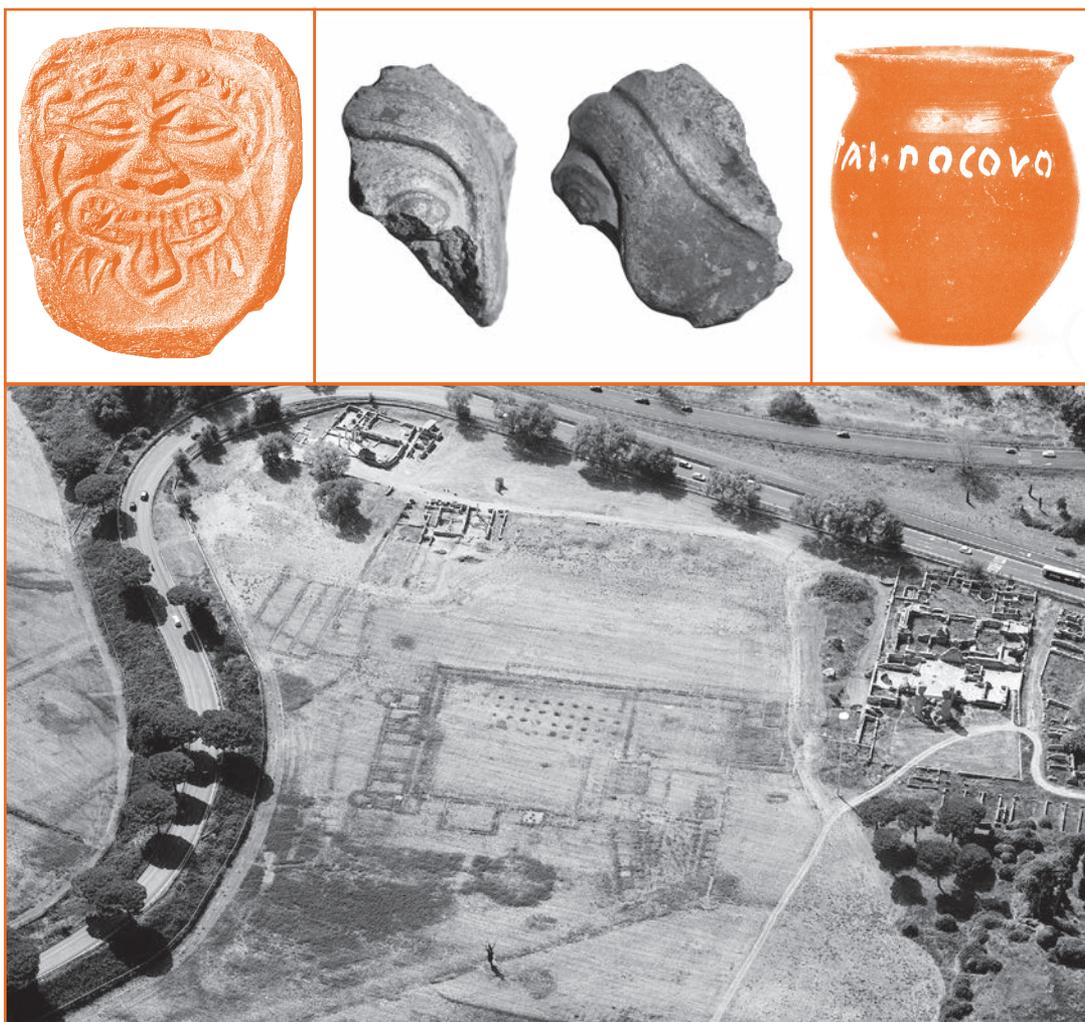


SAPIENZA UNIVERSITÀ DI ROMA
DIPARTIMENTO DI SCIENZE DELL'ANTICHITÀ

SCIENZE DELL'ANTICHITÀ

26.1 – 2020



EDIZIONI QUASAR

SAPIENZA UNIVERSITÀ DI ROMA
DIPARTIMENTO DI SCIENZE DELL'ANTICHITÀ

SCIENZE DELL'ANTICHITÀ

26 – 2020

Fascicolo 1

EDIZIONI QUASAR

DIPARTIMENTO DI SCIENZE DELL'ANTICHITÀ

Direttore
Giorgio Piras

Comitato di Direzione

Anna Maria Belardinelli, Carlo Giovanni Cereti, Cecilia Conati Barbaro, Maria Teresa D'Alessio, Giuseppe Lentini, Laura Maria Michetti, Francesca Romana Stasolla, Alessandra Ten, Pietro Vannicelli

Comitato scientifico

Graeme Barker (Cambridge), Martin Bentz (Bonn), Corinne Bonnet (Toulouse), Alain Bresson (Chicago), M. Luisa Catoni (Lucca), Alessandro Garcea (Paris-Sorbonne), Andrea Giardina (Pisa), Michael Heinzelmann (Köln), Mario Liverani (Roma), Paolo Matthiae (Roma), Athanasios Rizakis (Atene), Avinoam Shalem (Columbia University), Tesse Stek (Leiden), Guido Vannini (Firenze)

Redazione

Laura Maria Michetti
con la collaborazione di Martina Zinni

INDICE

I SEZIONE. ORIENTE E MEDITERRANEO	p.	1
C.G. Cereti – B. Faticoni – C. Marchetti – G. Terribili – M. Vassalli, <i>Il monumento di Narseh da Paikuli a Sulaimaniyah. Nuove attività archeologiche e allestimento espositivo del materiale presso lo Slemani Museum</i>		3
II SEZIONE. ROMA E IL LAZIO		13
M. Turci – M. Uehara – P.-E. Mathé, <i>Per un'ipotesi di identificazione del Foro di Aureliano e del praetorium di Ostia. Prime osservazioni a partire dalle indagini geofisiche</i>		15
L. Attenni – L. Ebanista – D. Nonnis, <i>La collezione Dionigi di Lanuvio: una nota su due iscrizioni vascolari d'età medio-repubblicana</i>		41
L.M. Michetti – M.C. Biella – B. Belelli Marchesini – F. Toppetti – D. Sarracino – M. Bischeri – M. Pacifici – L. Stomeo – L. Bianchi, <i>Novità dall'area urbana di Veio. Recenti indagini nel distretto meridionale</i>		61
F. Materazzi – M. Pacifici, <i>Novità dall'area urbana di Veio. Telerilevamento multispettrale da drone e indici di vegetazione: nuovi strumenti per l'identificazione dei crop-mark dall'area di Campetti</i>		95
M. Bonadies – M. Zinni, <i>La tomba "del tripode" di Falerii Veteres. Note su un contesto funerario di età arcaica</i>		119
I. Corda, <i>Diritto e società dei Sabini nelle fonti letterarie. Alcune osservazioni preliminari</i> ..		147
G.M. Annoscia, <i>Desinat hinc bellator atrox, iam desinat hostis: la guerra a Leopoli-Cencelle in epoca medievale</i>		157
S. Orlandi – G. Almagno – A. Mincuzzi, <i>EDR – effetti collaterali 6</i>		171
M.C. Capanna – C. di Fazio, <i>Tecnologia per le scienze umane. Il progetto ARCA e l'esplorazione automatica della conoscenza</i>		183
III SEZIONE. ITALIA MERIDIONALE		203
M. Galli – F. Coletti – A. Ciccola – I. Serafini, <i>Archeologia e archeometria del tessuto antico: un gruppo di manufatti aurei dall'area vesuviana (Pompei, Ercolano, Oplontis)</i>		205

MARIA CRISTINA CAPANNA – CLARA DI FAZIO

TECNOLOGIA PER LE SCIENZE UMANE.
IL PROGETTO ARCA E L'ESPLORAZIONE AUTOMATICA DELLA CONOSCENZA

ACADEMIC RESEARCH CREATIVITY ARCHIVES

Nato dalla cooperazione tra Sapienza Università di Roma, TSP (Tecnologia e Servizi Professionali)¹ e L'Erma di Bretschneider² e avviato nel 2019 grazie al programma di finanziamento della Regione Lazio “Creatività 2020”³, il progetto di ARCA (*Academic Research Creativity Archives*) vede i Dipartimenti di Scienze dell'Antichità e di Ingegneria Informatica, Automatica e Gestionale (DIAG)⁴ impegnati in stretta collaborazione scientifica nello sviluppo di una soluzione digitale innovativa per le ricerche semantiche e bibliografiche avanzate.

Il progetto, fondato sull'interdisciplinarietà, sull'integrazione e sulla condivisione di conoscenze e competenze tecnico-scientifiche tra *technologies* e *humanities*, suggerisce la combinazione tra i metodi di indagine propri delle discipline umanistiche e gli strumenti del mondo ICT (*Information and Communication Technologies*), ponendosi in linea con i principi dell'area “Industrie Creative e Digitali”, definita dalla Regione Lazio sulla base della *Smart Specialisation Strategy* dei territori elaborata su scala europea. Obiettivo di ARCA, infatti, è la creazione di una piattaforma integrata di prodotti editoriali e servizi digitali in grado di offrire nuove modalità ed esperienze di ricerca semantica e bibliografica, rinnovando il metodo tradizionale. Nell'accogliere i valori espressi nel programma europeo “Europe in a changing world - Inclusive, innovative and reflective societies”, l'attenzione dei ricercatori dei due dipartimenti coinvolti si concentra anche su un secondo, importante, aspetto, concernente la fruizione creativa dei contenuti accademici che saranno resi disponibili nella piattaforma ARCA, orientata verso più attuali forme di sviluppo e produzione multimediale, distribuzione di *cross-media* e post-produzione digitale.

¹ Azienda operante dal 2012 sul mercato ICT (*Information and Communication Technologies*) con soluzioni innovative per organizzazioni pubbliche e private, coniugando i principi della “Scienza dei Servizi” e del “Governo delle Informazioni”. La società TSP vanta una stabile collaborazione con l'Università degli Studi di Milano Bicocca e la Sapienza Università di Roma in progetti di ricerca e sviluppo, i cui risultati sono stati alla base della creazione anche di due *spin-off*: NEXTT LAB S.r.l. - *Networked Expertises Technologies and Tools* e MAP-S S.r.l. T.S.P.

² Storica casa editrice, fondata nel 1896, da sempre specializzata e impegnata nella pubblicazione e distribuzione su scala internazionale delle ricerche storiche, archeologiche e storico-artistiche.

³ Programma di finanziamento promosso nell'ambito del Fondo Europeo per lo Sviluppo Regionale (FESR) e del POR 2014-2020 (Programma Operativo Regione Lazio).

⁴ Le linee di ricerca del Dipartimento di Ingegneria Informatica Automatica e Gestionale investono la qualità dei dati, data *warehousing*, integrazione e scambio di dati, sistemi informativi basati su ontologie, *digital records management and preservation*, architetture a servizi, modellazione e realizzazione di servizi, sintesi e composizione di servizi, gestione di processi e di *workflow*, *data mining* e *machine learning* applicati agli *smart spaces*, interfacce utente, ingegneria dell'usabilità e dell'accessibilità, visualizzazione dell'informazione, modellazione, progettazione e realizzazione di *smart environments & communities*, applicazioni multimediali. Chi scrive è impegnato nell'*équipe* di lavoro, coordinata da Massimo Mecella per il DIAG, Paolo Carafa e Domenico Palombi per il DSA, con i colleghi Eleonora Bernasconi e Miguel Ceriani (DIAG) per l'analisi e lo sviluppo del progetto, Paolo de Paolis, Andrea Grazian, Elena Pavanello, Vincenza Scala (DSA) per la verifica e la validazione del prototipo. Ai risultati tecnico-scientifici di questa proficua esperienza di collaborazione e ricerca interdipartimentale sarà dedicato un contributo congiunto in via di elaborazione.

Nella prospettiva del “Restoring, preserving, valuing & managing the European Cultural Heritage (tangible & intangible)”, l’approccio intersettoriale di ARCA coniuga le istanze del mondo accademico e dell’editoria scientifica con le spinte dell’innovazione tecnologica e mira a far convergere le rinnovate esigenze di mercato dell’industria culturale sulle odierne proposte delle imprese ICT, in grado di mettere a disposizione ulteriori prodotti e strumenti al servizio della diffusione trasversale della cultura e della condivisione interdisciplinare del sapere, in un dialogo costante con le istituzioni culturali, le università e gli enti di ricerca.

M.C.C., C.d.F.

TECNOLOGIA E CONTESTO DI RIFERIMENTO

Il progetto ARCA è ispirato dalla volontà di promuovere la diffusione della ricerca scientifica in campo umanistico, valorizzandone i risultati e amplificando le possibilità di reperimento e fruizione dei prodotti e dei contenuti, tramite la creazione di un servizio che possa introdurre nel settore elementi e soluzioni derivanti dalle più moderne metodologie e innovazioni tecnologiche.

Da diversi anni il settore dell’editoria, su scala nazionale e internazionale, vive una fase di transizione di cui si colgono gli esiti nella spinta data all’innovazione e all’applicazione di tecnologie abilitanti, tali da trasformare l’approccio all’uso dei contenuti e alla gestione dei dati. Sempre più solido è stato il consolidamento dell’editoria digitale, con investimenti nella ricerca industriale e impiego di sistemi informatici appropriati a soddisfare le esigenze dei clienti/destinatari. Se dal *digital publishing* all’editoria elettronica il passo è stato breve, del tutto rivoluzionaria è stata la trasmissione dei prodotti editoriali digitali attraverso il *web*⁵. Un punto di svolta nelle forme di diffusione dei contenuti è stato tracciato dall’interattività con l’utente finale, attraverso opportune e innovative interfacce, in cui creatività e usabilità si fondono per migliorare la *customer experience*.

La necessità di reperire, raccogliere, analizzare e integrare i dati già noti accomuna ciascun ricercatore, che in ambito umanistico lavora, essenzialmente, con le risorse bibliografiche. Se le edizioni digitali rappresentano ad oggi uno strumento agevolante nella raccolta delle informazioni in confronto alle tradizionali modalità di consultazione dei repertori cartacei, la possibilità di individuare in tempo reale molteplici connessioni semantiche all’interno di un archivio digitale in costante aggiornamento implica una radicale trasformazione nell’esperienza di ricerca e nella fruizione dei contenuti, con notevoli vantaggi in termini di reperibilità e accessibilità dei testi, combinazione e relazione tra i dati, creazione di apparati tematici e bibliografici utili per le ricerche e condivisibili.

ARCA mira a creare uno strumento inedito nel campo delle scienze umane, e in particolar modo nelle discipline storico-archeologiche, innovando le modalità di ricerca bibliografica, che si adattano alle richieste degli utenti/ricercatori e alle trasformazioni intercorse nei modi di condivisione e diffusione dei dati e dei prodotti delle ricerche. Preliminare alla elaborazione del progetto è stata l’analisi delle esigenze dei ricercatori, primi destinatari della piattaforma, per i quali i “libri” costituiscono la risorsa fondamentale.

Nel sistema di ricerca bibliografica tradizionale, oltre agli schedari cartacei, si utilizzano interfacce dedicate per interrogare, attraverso parole chiave, un insieme strutturato di dati omogenei e controllati, ovvero il catalogo elettronico di una biblioteca.

L’*International Standard Bibliographic Description* (ISBD) indica quali siano le informazioni indispensabili nella catalogazione di ogni documento:

⁵ Sul tema delle biblioteche digitali vd. BIAGETTI 2019.

Titolo
Autore, Curatore, Traduttore
Edizione
Luogo di edizione
Editore
Data di pubblicazione
Descrizione fisica
Eventuali note
Numero ISBD/ISSN
Classificazione e soggettazione

Sono questi i campi interrogabili utilizzando il motore di ricerca del catalogo elettronico di una singola biblioteca, oppure impiegando le *Library Network*, oggi anche on-line e liberamente fruibili, che integrano le risorse elettroniche di più biblioteche. Tuttavia, è chiaro come tali modalità di ricerca risultino statiche e univoche e non consentano di correlare argomenti e ambiti tematici; senza contare la mancata visualizzazione del testo.

L'integrazione dei dati e la possibilità di interagire con sistemi più dinamici, in grado di "anticipare" le richieste dell'utente (*Recommendation Systems*), risulta essere una delle esigenze maggiormente avvertite anche in campo umanistico, in aggiunta alle necessità di accesso simultaneo a una pluralità di informazioni, che consentano di reperire in parallelo il maggior numero di risorse bibliografiche sul tema di interesse e su ambiti interdisciplinari correlati. Al pari dei consueti metodi di ricerca bibliografica (sia cartacea sia elettronica), non risultano dinamici, né automatici o replicabili, il processo di elaborazione delle associazioni tematiche e la creazione di insiemi bibliografici basati sulla concatenazione di singoli concetti. Un altro aspetto problematico riguarda, infine, la dispersione e la frammentazione delle risorse digitali disponibili. Nel tentativo di superare questa segmentazione di contenuti e documenti, il progetto ARCA suggerisce l'uso di un sistema complesso e integrato che garantisca l'interoperabilità tra diversi livelli di approfondimento (articolati secondo il profilo e la richiesta dell'utente). ARCA, dunque, si pone nell'ottica di una soluzione più olistica negli esiti delle ricerche e nella reperibilità dei prodotti editoriali.

Per realizzare questi obiettivi, ARCA abbandona la ricerca tradizionale basata sulla presenza/assenza di parole chiave univoche in campi dedicati, in favore di una ricerca integrale nel corpo del testo digitalizzato, o nato in formato digitale. Il sistema, estraendo le entità e proponendo associazioni semantiche e connessioni ipertestuali, moltiplica esponenzialmente i risultati e trasforma l'esperienza di consultazione in una nuova modalità di accesso, fruizione e acquisizione dei contenuti.

Partendo dalla indicizzazione e archiviazione dei metadati e dalla digitalizzazione idonea alla codificazione semantica di un segmento del patrimonio editoriale della casa editrice *partner*, la piattaforma ARCA ha generato una "base di conoscenza integrata", realizzata mediante l'uso di una tecnologia abilitante definita *Knowledge Graph* (KG)⁶. Al sistema di consultazione si affiancano servizi a supporto dell'utente (sia specialista sia *general public*) intenzionato ad esplorare, a distinti livelli di approfondimento, la base di conoscenza.

A livello tecnico-scientifico l'idea progettuale di ARCA si fonda sui risultati di una precedente esperienza di ricerca in campo informatico: il progetto DaCENA (*Data ContExt for News Articles*)⁷, sviluppato dal Dipartimento di Informatica, Sistemistica e Comunicazione dell'Uni-

⁶ Vd. il contributo di dettaglio in CERIANI *et al.* 2020, pp. 75-81, con rif.

⁷ Per ulteriori riferimenti: <http://www.densitydesign.org/research/dacena/>. Per la presentazione dei risultati scientifici: PALMONARI *et al.* 2015, pp. 133-137; BIANCHI *et al.* 2017, pp. 120-135.

versità degli Studi di Milano Bicocca in collaborazione con il Dipartimento di Design del Politecnico di Milano. In estrema sintesi, DaCENA costituisce una applicazione *web* rivolta all'ambito giornalistico e creata con l'intento di fornire all'utente impegnato nella lettura di un articolo di giornale ulteriori informazioni di contesto sul tema trattato, ricavandole da un *Knowledge Graph*. L'applicativo consente di estrarre dal testo le "entità" collegate al *KG*, a loro volta impiegate nella costruzione di altre associazioni semantiche che ampliano la raccolta, qualitativa e quantitativa, delle informazioni disponibili, definendo così il contesto di riferimento dell'articolo e dilatando l'orizzonte conoscitivo. Grazie all'impiego di tecniche specifiche, un elemento interessante e innovativo introdotto da DaCENA risiede nella presentazione all'utente non solo delle relazioni più rilevanti sull'argomento in oggetto, ma anche di quelle che potrebbero risultare "inaspettate", accrescendo in tal modo l'effetto *serendipity* nell'utente. Questo aspetto identifica DaCENA con un *contextual exploration tool*, applicato al campo del *relational Data Journalism* e composto da un *Text & Data Analyzer* e da un *Contextual Explorer*, rispettivamente lo strumento tecnico per il recupero delle informazioni di contesto come associazioni semantiche e l'interfaccia utente.

Fermo restando il punto di riferimento fissato dal prototipo tecnico di DaCENA, la sfida di ARCA consiste nello sviluppo dell'applicativo in un ambito quale l'editoria umanistica (e in particolare archeologica nella fase di sperimentazione), e nell'integrazione di tecnologie che possano mettere a disposizione dell'utente ulteriori prodotti e servizi, con adeguati livelli di *performance* e aggiornate modalità di fruizione digitale dei contenuti⁸, non escluse quelle audiovisive e basate sullo *storytelling* (vd. *infra*) Su questo fronte i colleghi ricercatori del Dipartimento di Ingegneria Informatica Automatica e Gestionale stanno sviluppando appositi *crawler* specializzati nel *multimedia* e nel *social signal processing*, finalizzati alla raccolta e all'analisi del materiale disponibile, coerentemente associato al grafo semantico utilizzato per i documenti interni. La combinazione semantica dei contenuti, e delle preferenze dell'utente, si avvarrà dell'applicazione di tecniche di *matchmaking* basato su *Artificial Intelligence* (AI)⁹, che consentiranno di personalizzare l'organizzazione dei risultati.

L'ecosistema digitale in *cloud* rappresentato da ARCA si inserisce, con una proposta informatica e tecnologica inedita in ambito umanistico, nell'odierno scenario di trasformazione del mercato editoriale, in particolare dell'editoria scientifica. ARCA può costituire un "vantaggio competitivo potenziale" per coloro che, a diversi livelli, operano nel settore, ampliandone le edizioni digitali e di conseguenza le opportunità di commercializzazione dei prodotti pubblicati. Al contempo, la possibilità di distribuzione secondo il modello SaaS (*Software as a Service*) consente anche un incremento dell'attuale offerta del mondo ICT nel mercato dell'industria culturale.

In sintesi, i servizi applicativi della piattaforma ARCA sono rappresentati da:

- Digitalizzazione di contenuti e prodotti editoriali (libri, riviste, articoli, miscellanee, cataloghi)
- Formalizzazione semantica della conoscenza estratta dai prodotti editoriali
- Creazione di una base di conoscenza integrata (*Knowledge Base*)
- Esplorazione, a diversi livelli e a seconda del profilo utente, della base di conoscenza
- Ricerca semantica avanzata all'interno della base di conoscenza (KG)

⁸ In tal senso il progetto ARCA si inquadra nel piano di azione approvato dalla Agenda Digitale Lazio, sulla scorta della più ampia Agenda Digitale Europea, e interviene nell'ambito delle cosiddette "Comunità Intelligenti" proponendo un processo di digitalizzazione, ricerca semantica e pubblicazione di contenuti *on-demand*.

⁹ Vd. BERARDI *et al.* 2003; MARRELLA *et al.* 2017, pp. 1-43.

- *Ranking* avanzato per una selezione intelligente delle informazioni estratte dalla base di conoscenza
- Accesso ai contenuti editoriali e acquisizione secondo gli interessi

Si tratta di uno strumento innovativo di grande utilità in ambito accademico, tale da favorire lo scambio e la condivisione su scala internazionale delle pubblicazioni dei risultati di studi e ricerche nelle diverse discipline. Il prototipo della piattaforma sarà infatti fruibile sia in lingua italiana sia in inglese. Rispondendo all'esigenza dei ricercatori di esplorare basi di conoscenza nel contesto di riferimento e di comparare i risultati, ARCA, mediante un sistema dinamico e interoperabile, restituisce collegamenti integrati, senza escludere la possibilità di connessioni inaspettate, tali da generare nuove idee. Le entità estratte dai testi in .pdf forniti dalla casa editrice sono collegate a un *Knowledge Graph* (KG) esterno, in particolare al KG del *dataset* DBpedia, che restituisce le principali informazioni tratte dal *web*. Le informazioni esterne (DBpedia) e interne (metadati ed entità estratte dai testi in .pdf *in batch*) confluiscono, tramite l'uso di un *triple store*, in un KG appositamente realizzato: *ARCA Knowledge Graph*, ovvero la base di conoscenza organizzata sotto forma di concetti, esplorabile in ogni sua diramazione (Fig. 1). Oltre ad effettuare ricerche sull'intera libreria a disposizione, rintracciare e visualizzare i contenuti, l'utente potrà organizzare, gestire e depositare documenti e dati, creare note e liste bibliografiche ragionate, avviare le procedure di acquisto dei testi di interesse.

Per l'estrazione delle entità ARCA utilizza *Dandelion*¹⁰, l'API (*Application Programming Interface*) di Spazio Dati in grado di elaborare una classifica delle entità più ricorrenti e potenzialmente più significative per il testo. La ricerca di contenuti si avvale di due *tool*:

- *Entity Search*: ricerca un termine all'interno del database ed estrae tutte le connessioni che si diramano dall'entità principale
- *Road Map*: traccia i collegamenti tra le relazioni di due entità, offrendo la possibilità di sperimentare percorsi alternativi di ricerca e di navigare tra le associazioni semantiche, con un potenziale effetto *serendipity*

Ciascun nodo del KG è caratterizzato da pulsanti che consentono l'esplorazione e l'interrogazione del motore di ricerca:

- “Esplora la biblioteca”: per accedere all'elenco dei libri nella biblioteca virtuale
- “Esplora il grafo”: per accedere simultaneamente a una pluralità di informazioni articolate su diversi livelli di approfondimento
- “Centra e Riparti”: per avviare una nuova esplorazione, interagendo con l'espansione e la moltiplicazione delle informazioni restituite dalla *RoadMap*

In aggiunta al *Knowledge Graph* (KG) e a *React*, una libreria JavaScript per creare interfacce con visualizzazioni dinamiche e interattive, le tecnologie abilitanti impiegate da ARCA si riconoscono nei processi di *Named Entity Recognition* (NER), attività collegata all'estrazione delle informazioni e finalizzata a individuare e classificare le entità presenti nel testo secondo categorie predefinite, di *Named Entity Linking* (NEL), associazione tra le entità individuate mediante NER con entità del KG di riferimento; infine di *Natural Language Processing* (NLP),

¹⁰ Lanciata da Spazio Dati al “Pioneers Festival” di Vienna nel 2013, dandelion.eu fu selezionata tra le migliori 16 *startup* al mondo. Si tratta di un KG, collegato a motore semantico (*dataTXT*) per l'analisi automatica dei testi, con milioni di informazioni in formato *machine readable* su entità come aziende, persone, luoghi, eventi e concetti.

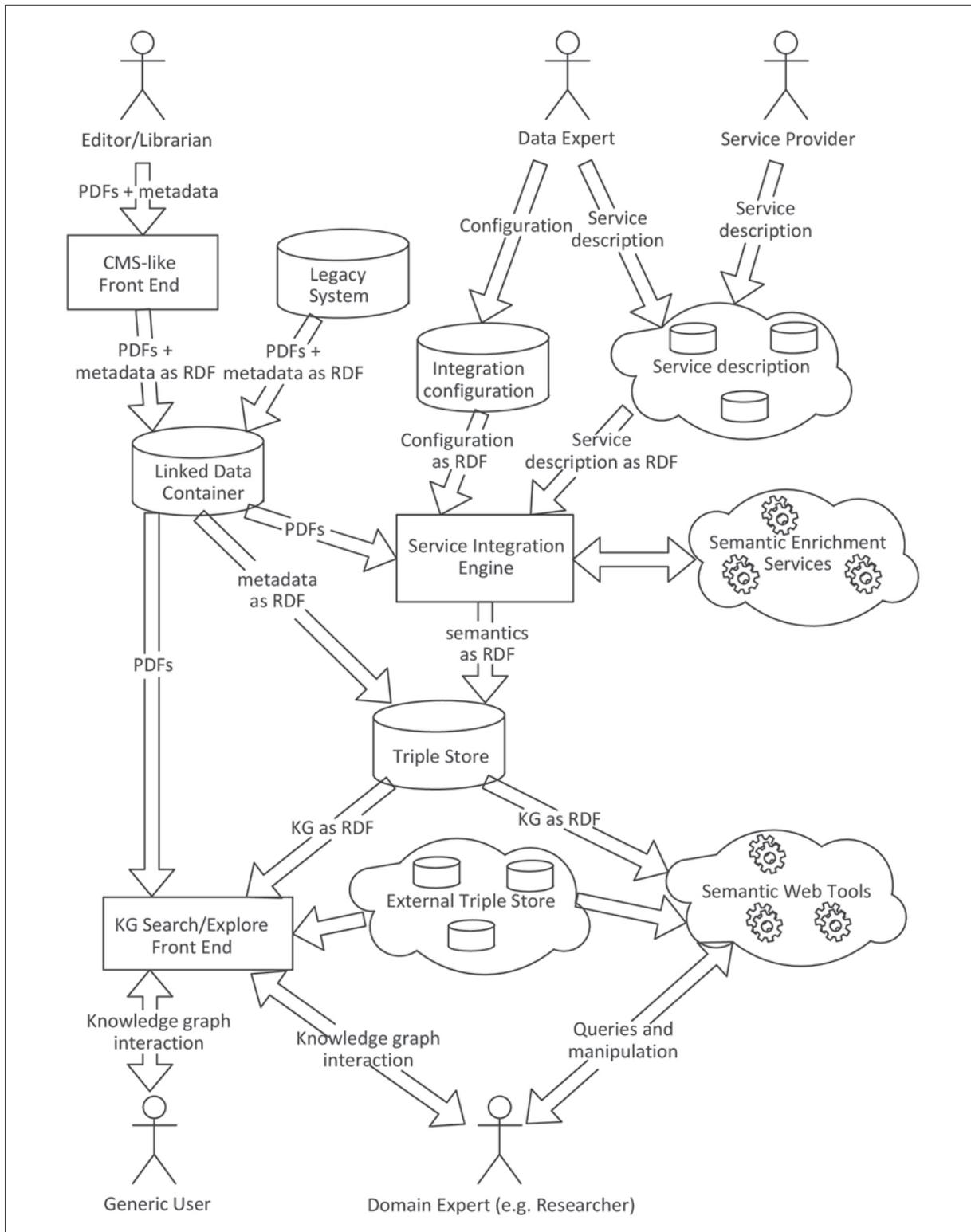


Fig. 1 – Pipeline ottimizzata per l'esplorazione semantica di una libreria (rielab. da CERIANI - BERNASCONI - MECELLA 2020: https://www.researchgate.net/publication/338986388_A_Streamlined_Pipeline_to_Enable_the_Semantic_Exploration_of_a_Bookstorepdf).

interessante campo di ricerca interdisciplinare tra informatica, intelligenza artificiale e linguistica, che mira allo sviluppo di algoritmi in grado di analizzare, rappresentare e “comprendere” il

linguaggio naturale, scritto o parlato¹¹. Il processo di NLP consiste in una profonda analisi del testo, che passa attraverso la restituzione tanto delle strutture e dipendenze sintattiche quanto delle categorie lessicali, la comprensione delle relazioni semantiche e la contestualizzazione del significato delle parole.

C.d.F.

CONTESTO BIBLIOGRAFICO E AMBITI SEMANTICI DELLA RICERCA

Trattandosi di un progetto sperimentale, nella fase preliminare, è stato selezionato, all'interno dell'offerta editoriale del *partner* L'Erma di Bretschneider, un nucleo di testi tratti dalla letteratura archeologica, storica e storico-artistica che avesse un comune denominatore. Tale comune denominatore è stato individuato nel Lazio, antico, medievale e moderno, per alcune sue caratteristiche:

- Ampiezza geografica
- Omogeneità culturale
- Fenomeni di lunga durata
- Centralità politica e culturale nell'ambito italiano ed europeo

Tale centralità si evince anche da semplici fattori numerici. *Latium* è il nome di regione antica tra le più citate nelle fonti letterarie antiche, annalistiche e non. Interrogando la *Perseus Digital Library*¹², inserendo *Latium* come parola chiave si ottengono 194 risultati tra i testi in lingua latina e 10 nei testi in greco in traduzione inglese; inserendo invece la parola *Sabina*, regione confinante con il *Latium vetus*, se ne ottengono 30 tra i testi in lingua latina e 1 solo nei testi in greco in traduzione inglese.

La varietà dei luoghi e delle popolazioni dovuta all'ampiezza e alla lunga durata di occupazione della regione offre innumerevoli temi di ricerca, nell'ambito dello studio della cultura materiale, della produzione artistica, della storia, della storia sociale ed economica o giuridica, della religione, e rende il "Lazio antico" tra i soggetti più studiati da storici, storici dell'arte e archeologi.

Interrogando i cataloghi on-line delle biblioteche dei principali istituti/enti di ricerca nel campo umanistico risultano circa 850 i titoli in lingua italiana che contengono la parola Lazio, cui vanno aggiunti altri 440 titoli in altra lingua che contengono la parola *Latium*. Meno di 300 sono, complessivamente, i titoli che contengono la parola Sabina.

Il campione scelto per la sperimentazione è formato da circa 200 volumi tra monografie, edizioni periodiche, riviste, atti di convegno e studi miscelanei. L'ampio dominio di produzione specialistica costituisce la base di conoscenza rappresentativa, valido modello d'insieme per la sperimentazione in corso.

Il ruolo principale del gruppo di ricerca del Dipartimento di Scienze dell'Antichità è quello di verificare che il prodotto risponda alle reali necessità degli utenti finali, garantendo la correttezza dei criteri di connessione e la coerenza scientifica delle ontologie di riferimento nei risultati delle ricerche.

Per definire gli ambiti semantici e valutare l'esattezza delle ontologie del KG di riferimento di DBpedia si ritiene che la struttura logica adottata dal Progetto Dyabola sia il migliore e più idoneo

¹¹ Sul ruolo e sul fondamentale apporto degli studi cognitivi, della linguistica e delle ricerche umanistiche nello sviluppo della *computer science* vd. BOD 2013; ID. 2019. Sul NLP BIRD *et al.* 2009; MAYNARD *et al.* 2017.

¹² <http://www.perseus.tufts.edu/hopper/>.

termine di confronto per il settore umanistico, data la sua completezza e la sua articolazione in termini tipologici, cronologici e topografici, aspetti fondamentali delle ricerche storico-archeologiche. Il software per la creazione e la navigazione tra dati bibliografici e collezioni di immagini è stato realizzato dalla casa editrice tedesca Biering & Brinkmann per l'Istituto Archeologico Germanico che possiede una delle più ampie collezioni librerie di argomento storico-archeologico, garanzia del fatto che lo strumento sia stato progettato prevedendo la più ampia casistica di soggetti. Il progetto di catalogazione digitale, intrapreso negli anni Novanta, deriva da quello di catalogazione tradizionale, avviato alla fine degli anni Cinquanta dagli studiosi dell'Istituto, primi fruitori dello strumento, e poi proseguito e perfezionato da un reparto dedicato dello stesso Istituto, costituito da personale qualificato¹³. Sono 34 i rami principali dell'albero dei soggetti; l'indice generale conta 3.388.000 voci di soggetto¹⁴, con una varietà idonea per effettuare ricerche in ambito storico-archeologico su vasta scala geografica e ampio arco cronologico.

M.C.C.

ARCA: ATTIVITÀ, ANALISI E SPERIMENTAZIONE PRELIMINARE

I due enti di ricerca coinvolti nel progetto, i Dipartimenti di Scienze dell'Antichità e di Ingegneria Informatica, Automatica e Gestionale, cooperano in stretta relazione. Nella prima fase del lavoro i ricercatori di archeologia hanno illustrato le necessità dell'utenza e gli ingegneri le potenzialità dello strumento; nella fase operativa i secondi hanno sviluppato i componenti della piattaforma e i primi li hanno sottoposti a validazione. Le annotazioni e le risultanze sono utilizzate per risolvere le criticità e migliorare le prestazioni del sistema.

Le principali attività che il gruppo di ricerca del Dipartimento di Scienze dell'Antichità ha individuato come necessarie per la realizzazione dello strumento sono:

- Verifica della completezza e la correttezza del *Knowledge Graph* di DBpedia, utilizzato da DaCENA per l'estrazione delle entità
- Identificazione di ulteriori librerie e di ontologie liberamente fruibili per il *web* semantico da affiancare a quelle di DBpedia per ampliare la base di conoscenza
- Progettazione di nuove tassonomie specialistiche nel campo delle ricerche archeologiche e definizione delle nuove entità da affiancare a quelle di DBpedia
- Verifica delle entità estratte dal sistema sul campione di pdf selezionati, l'identificazione delle criticità e la ricerca di possibili azioni correttive (*Tab. 1*)

Verifica della completezza e correttezza del Knowledge Graph di riferimento DBpedia.

L'albero tassonomico di DBpedia¹⁵ conta 4.233.000 istanze articolate in 5 classi principali:

Place	735000
Person	1450000
Work	411000
Species	251000
Organisation	241000

¹³ Per la storia della Progetto Dyabola si consulti la pagina http://www.dyabola.de/en/indexfrm.htm?page=http://www.dyabola.de/en/projects/detail/sachkatalog_145.htm

¹⁴ Dato aggiornato ad Aprile 2020.

¹⁵ <http://mappings.dbpedia.org/server/ontology/classes/>

CRITICITÀ	POSSIBILI AZIONI CORRETTIVE	
	Fase I	Fase II
Errata o mancata disambiguazione		Sviluppo di soluzioni (deep and machine learning) in campo NLP* * <i>Eventuali sviluppi futuri del progetto</i>
Errata lettura dell'OCR		Sviluppo di tecniche di correzione post OCR* * <i>Eventuali sviluppi futuri del progetto</i>
Errata interpretazione (e conseguente attribuzione di entità non pertinenti)		Sviluppo di soluzioni in campo NER, NEL* * <i>Eventuali sviluppi futuri del progetto</i>
Errata o mancata lettura di altre lingue (e conseguente attribuzione di entità non pertinenti)	Integrazione con vocabolari	
Errata o mancata lettura di nomi propri (e conseguente attribuzione di nomi non pertinenti)		Sviluppo di soluzioni (deep and machine learning) in campo NLP* * <i>Eventuali sviluppi futuri del progetto</i>
Errata o mancata lettura di locuzioni (e conseguente attribuzione di entità non pertinenti)	Per i luoghi di Roma antica integrazione degli indici del <i>LTUR</i> Integrazione con ontologie dedicate	
Errata disambiguazione di verbi, aggettivi, sostantivi, sinonimi (e conseguente attribuzione di entità non pertinenti)	Integrazione con vocabolari	Integrazione con modelli di reasoning* * <i>Eventuali sviluppi futuri del progetto</i>
Estrapolazioni di città e concetti dalla bibliografia e dalle note a piè di pagina	Applicazione di filtri di esclusione	
Associazione (errata) dei numeri romani (I, II, III ...IX...) ai secoli		Integrazione con modelli di reasoning* * <i>Eventuali sviluppi futuri del progetto</i>
Errato riconoscimento di Abbreviazioni Bibliografiche e attribuzioni a concetti/entità non pertinenti	Affiancare alle librerie di DBpedia gli elenchi di abbreviazioni bibliografiche di prassi utilizzate nelle edizioni scientifiche di argomento archeologico/umanistico: sistema francese (Année Philologique), tedesco (Archäologischer Anzeiger); anglosassone (rete URBiS)	
Errato riconoscimento di abbreviazioni e sigle Archivistiche e attribuzioni a concetti/entità non pertinenti	Affiancare alle librerie di DBpedia gli elenchi di abbreviazioni archivistiche comunemente utilizzate nelle edizioni scientifiche di argomento archeologico/umanistico Integrazione con ontologie dedicate	
Errato riconoscimento di sigle e abbreviazioni editoriali e attribuzioni a concetti/entità non pertinenti	Applicazione di filtri di esclusione	
Fraintendimenti semantici		Sviluppo di soluzioni (deep and machine learning) in campo NLP* * <i>Eventuali sviluppi futuri del progetto</i>
Associazioni arbitrarie		Sviluppo di soluzioni in campo NER, NEL* * <i>Eventuali sviluppi futuri del progetto</i>
Estrapolazione da iscrizioni trascritte		Sviluppo di tecniche di correzione post OCR* * <i>Eventuali sviluppi futuri del progetto</i>

Tab. 1 – Criticità individuate e possibili azioni correttive.

Trattandosi di un prodotto finalizzato ad uno specifico contesto di studi umanistici, si ritiene utile validare prioritariamente le categorie collegate all'ambito disciplinare scelto, partendo dalle 3500 circa voci dell'albero di Dyabola (vedi *supra*); verificata la presenza, se ne valuterà anche la congruità del contenuto.

Identificazione di librerie liberamente fruibili e di ontologie da affiancare a DBpedia.

Data la specializzazione dell'ambito di applicazione della piattaforma e l'alto grado di specializzazione dell'utenza finale, è necessario l'uso di ontologie specifiche che permettano di ampliare il KG di riferimento in modo da raffinare l'estrazione delle entità. Nella prima fase di lavoro sono state selezionate alcune ontologie prodotte da enti di ricerca scientifica che garantiscono alla piattaforma correttezza e congruità dei contenuti. Sarà compito del gruppo di ricerca del Dipartimento di Ingegneria Informatica, Automatica e Gestionale, in una successiva fase della ricerca sperimentale, identificare le soluzioni tecniche e tecnologiche per affiancare tali librerie a quelle del KG di riferimento.

Per la geografia storica:

- Pleiades, creato da Ancient World Mapping Center, Institute for the Study of the Ancient World, Stoa Consortium, <https://pleiades.stoa.org/>
- Digital Atlas of the Roman Empire (DARE), creato da Department of Archaeology and Ancient History, Lund University, Sweden; <http://dare.ht.lu.se/>

Per i personaggi storici:

- SNAP Graph, Progetto: Standards for Networking Ancient Prosopographies: Data and Relations in Greco-Roman Names; <http://snapdrgn.net/database>

Per l'epigrafia:

- Epigraphic Database Heidelberg (EDH), creato da Seminar für Alte Geschichte und Epigraphik, Universität Heidelberg, <https://edh-www.adw.uni-heidelberg.de/home>

Per le fonti letterarie antiche:

- Perseus Classical Library creato dall'Istituto Archeologico Germanico <http://archwort.dainst.org/de/vocab/>

Verifica delle entità estratte dal sistema; identificazione delle criticità e possibili azioni correttive.

Tre sono i percorsi ideati per sottoporre a verifica le entità estratte dal sistema: 1) la ricerca "tradizionale" di parole/concetti nei testi e controllo della presenza/assenza degli stessi tra le entità estratte dal sistema; 2) la verifica sistematica delle entità estratte dal sistema; 3) la lettura integrale di alcuni dei testi da parte dei ricercatori per l'individuazione "manuale" di tutte le entità e confronto con quelle estratte dal sistema.

1) Sono stati scelti ipotetici temi di ricerca, alcuni più ampi, e altri più specifici. Si è proceduto "manualmente" ad effettuare la ricerca sul campione limitato dei testi selezionati per la sperimentazione¹⁶. Le parole chiave (declinate al singolare e al plurale) insieme a quelle dello stesso

¹⁶ L'operazione di estrazione delle entità dai testi non è banale e nella fase iniziale di sperimentazione sono state effettuate diverse calibrizioni dei parametri di elaborazione per identificare la soluzione che permetta di ottenere i risultati più soddisfacenti. Questo ha implicato la necessità di ripetere l'operazione di estrazione più volte su uno stesso testo, variando tali parametri. È stato valutato che un campione di testi pari al 10% del totale è sufficiente per ottenere buoni risultati in un tempo adeguato che includa anche la verifica dei risultati e la successiva fase di perfezionamento

ambito semantico sono state cercate tradizionalmente nei pdf. Il risultato è stato confrontato con le entità estratte da ARCA.

Un esempio.

- a. La villa (intesa come edificio romano, destinato all'abitazione e alla attività agricola, isolato nella campagna).
- b. La gioielleria antica (per tale tema la ricerca è stata effettuata su le parole chiave del gruppo semantico dei gioielli: anello, collana, orecchino, pendente, fibula, ecc.)

La ricerca manuale del termine villa in uno dei testi ha portato ai seguenti risultati:

- a. *Villa publica* (edificio pubblico di Roma): 2 volte;
Villa Albani (residenza nobiliare Settecentesca sulla via Salaria): 1 volta;
Villa (intesa come antica residenza nella campana, in particolare la villa tuscolana di Cicero-
ne): 2 volte
- b. La ricerca manuale relativa alla gioielleria antica non ha dato risultati.

Nelle entità estratte dal sistema il risultato coincide perfettamente solo quando sono applicati alcuni dei parametri di estrazione.

La verifica è stata analogamente condotta con altri temi di ricerca conseguendo gli stessi risultati che hanno portato a identificare i parametri migliori da applicare alle successive estrazioni delle entità.

2) Effettuata l'estrazione, l'algoritmo categorizza le entità in gradi di rilevanza, in base alla ricorrenza nel testo. Tre sono le categorie individuate: i "top-concepts", ovvero le dieci entità più ricorrenti in ciascun contributo; i "concepts", ovvero le entità ricorrenti più di 20 volte; gli "other concepts", le entità al di sotto della precedente soglia (*Fig. 3*).

Sono state sottoposte a validazione tutti i "top-concepts" e i "concepts" estratti dal sistema e si sta procedendo con la verifica delle "altre entità" su 6 dei pdf selezionati. In quasi tutti i casi, e indifferentemente dai parametri utilizzati, risultano corretti i "top-concepts" e i "concepts", mentre la maggior parte delle errate estrazioni si concentrano nelle entità di minore rilevanza. In tutti i casi, gli errori sono riconducibili a poche tipologie, per le quali si stanno già adottando le prime soluzioni o per le quali è stata identificata una soluzione che potrà essere adottata in una successiva fase di sperimentazione.

- Errato scioglimento di sigle e abbreviazioni. Tale tipo di errore può essere risolto affiancando alle ontologie esistenti quelle specifiche del settore (abbreviazioni delle riviste scientifiche, secondo i sistemi bibliografici comunemente adottati nel mondo accademico, corpora, lessici, archivi e fondi dedicati), eventualmente imponendo una priorità di scelta su queste piuttosto che su quelle del KG di DBpedia;
- L'attribuzione di parole latine (o di altre lingue) ricondotte all'inglese o all'italiano. Tale tipo di errore può essere risolto affiancando i dizionari delle due lingue antiche specifiche del campo di studi umanistici (Latino e Greco) e delle principali lingue europee (Inglese, Italiano, Francese, Tedesco), forse ipotizzando un filtro con selezione delle lingue da prendere in considerazione in fase di settaggio dei parametri di estrazione delle entità.
- Errata disambiguazione delle entità per assenza nel KG di riferimento dell'entità corretta. Ad esempio, il nome antico greco *Arcesilaus* fu sia di un filosofo greco del IV-III

della piattaforma, la ripetizione del processo e la valutazione per il collaudo finale. Per gli aspetti tecnici e i parametri di estrazione vd. ora CERIANI *et al.* 2020, pp. 75-81.

secolo a.C. che di tre scultori (di V, IV e I secolo di I a.C.). L'unico personaggio presente nell'ontologia di riferimento di DBpedia è il primo ed è stato quindi impossibile riconoscere lo scultore di I secolo a.C. Questo tipo di errore, come tutti quelli dovuti a lacune nelle ontologie, sono risolvibili affiancandone di specifiche, in parte già individuate, o creando nuove ontologie da inserire nella struttura di DBpedia implementabile dagli utenti registrati.

- Errata lettura dell'OCR (Riconoscimento Ottico dei Caratteri). È stato implementato nella piattaforma un algoritmo per ottimizzare la lettura del pdf, ma è stato verificato che tale "ottimizzazione" risolve solo alcuni problemi, generandone altri di maggiore entità. Per la soluzione di tale problema deve essere dedicato un ulteriore modulo di ricerca.

- Errata "categorizzazione" delle entità. Anche quando formalmente corrette, le entità vengono inserite nelle tre categorie in base ad un mero calcolo delle occorrenze, metodo che non sempre corrisponde a quello di uno studioso. Ad esempio, nel caso di un volume che ha per oggetto un quartiere dall'antica città di Roma, "Roma" dovrebbe in ogni caso risultare tra i "top-concepts", cosa che avviene solo in base al numero di occorrenze molto alto. Il problema potrebbe essere risolto imponendo alla piattaforma di "promuovere" tra i "top-concepts" quelli individuati automaticamente e che coincidono con le parole chiave e i soggetti presenti tra i metadati dei volumi (soggetto/i principale; periodo cronologico; area geografica di interesse/luogo principale)

- Errata disambiguazione tra due o più entità tra cui il sistema sceglie secondo parametri non ancora identificati. Si tratta, al momento, della tipologia di più difficile soluzione. Sarebbe necessario un avanzamento del progetto di ricerca verso lo sviluppo di soluzioni (*deep and machine learning*) in campo NLP.

In generale, soprattutto per i problemi di disambiguazione e di gerarchizzazione, sembra che il sistema abbia più difficoltà a identificare concetti astratti rispetto a quelli concreti. Questo vale tanto nel ricondurre le singole entità ad una gerarchicamente superiore (come il già citato esempio della città in cui si trovano gli edifici citati), quanto a cogliere il senso di allegorie e metafore: ad esempio la parola "arma" è sempre ricollegata all'oggetto materiale; o la parola "atmosfera" sempre rimanda all'entità concreta dell'"atmosfera terrestre" e non alla "condizione psicologica in un determinato ambiente". Ancora più ampio il divario tra entità estratta dal sistema – Pioggia (precipitazione atmosferica), dalla locuzione "acqua piovana" citata nel testo quale elemento da smaltire, e che riconduce all'entità "idraulica", "opera idraulica" "ingegneria idraulica". In tutti questi casi, si propone, allo stato attuale, di imporre al sistema di NON disambiguare, lasciando all'utente la possibilità di scegliere l'entità corretta.

3) La verifica tramite il confronto tra le entità estratte dalla lettura integrale dei testi e quelle estratte dal sistema ha evidenziato due elementi fondamentali: da una parte la maggiore capacità del ricercatore di individuare concetti astratti rispetto a quanto non possa fare il sistema che parte dalla base oggettiva delle parole che formano il testo; d'altra parte, il sistema, oggettivo e obiettivo, estrae entità anche lontane dall'argomento principale del testo, che lo studioso specialista tende, inconsapevolmente, a non cogliere. Grazie a questa "oggettività" da ciascun testo possono essere estratte molte più entità semantiche. Ciò permette la creazione di innumerevoli relazioni tra i volumi e nella base di conoscenza, aprendo scenari inaspettati per la ricerca e offrendo un notevole vantaggio nella diffusione dei contenuti editoriali.

ARCA: SEMANTIC WEB, SERENDIPITY E DIGITAL STORYTELLING

When, Where, What, Who, Why? La nota regola giornalistica delle 5W resta valida ed efficace per ogni costruzione narrativa, compreso lo *storytelling*. Affinché questa forma di comunicazione strategica risulti incisiva è necessario che l'impianto narrativo sia uniforme e congruo rispetto ai messaggi da veicolare e ai destinatari da raggiungere. Se conoscere a fondo l'oggetto della "narrazione" è la chiave di partenza, assolutamente fondamentale è tenere a mente il carattere distintivo del target di riferimento, al quale devono adeguarsi il tono e i contenuti. Per catturare l'attenzione e suscitare l'interesse, la struttura di base dello *storytelling* può arricchirsi di dettagli, di ulteriori informazioni e di elementi utili, che devono però sempre risultare coerenti con le caratteristiche del tema, gli interessi e il livello degli utenti. Verificabilità, credibilità e controllo delle fonti risultano indispensabili per il successo della narrazione, che per essere pienamente efficace deve conservare una struttura logica e intellegibile. Il punto di forza di questa tecnica comunicativa risiede nella possibilità di trasferire dati, idee e suggerimenti in maniera diretta e coinvolgente. Particolarmente utili allo svolgimento del racconto sono le immagini e i video, che amplificano l'effetto sul destinatario. In un mondo in cui l'informazione tende a sovrastare l'utente, soltanto una "storia" realmente interessante, e ben fondata, può instaurare una relazione fiduciaria e creare un bisogno necessario per la fidelizzazione del lettore.

Queste coordinate generali valgono anche nel caso del *digital storytelling* che, mediante l'articolazione di contenuti *web* (video, audio, testi, immagini, mappe) organizzati in un sistema coerente, riesce a veicolare le informazioni in maniera interattiva, creando percorsi molteplici e relazioni analogiche o semantiche, tali da offrire all'utente una esperienza di ricerca stimolante.

In quest'ottica il sistema ARCA pone grande attenzione alla *user experience* fornendo strumenti intuitivi e modellati sui *desiderata* dell'utente e offrendo una varietà di contenuti e di dati che, eventualmente, possono condurre gli esiti delle ricerche verso ambiti inaspettati, ma di interesse e tali da innescare un soddisfacente "effetto sorpresa" (Fig. 2). Al contempo, i contenuti testuali e multimediali raccolti possono essere rielaborati e impiegati dagli utenti per costruire non

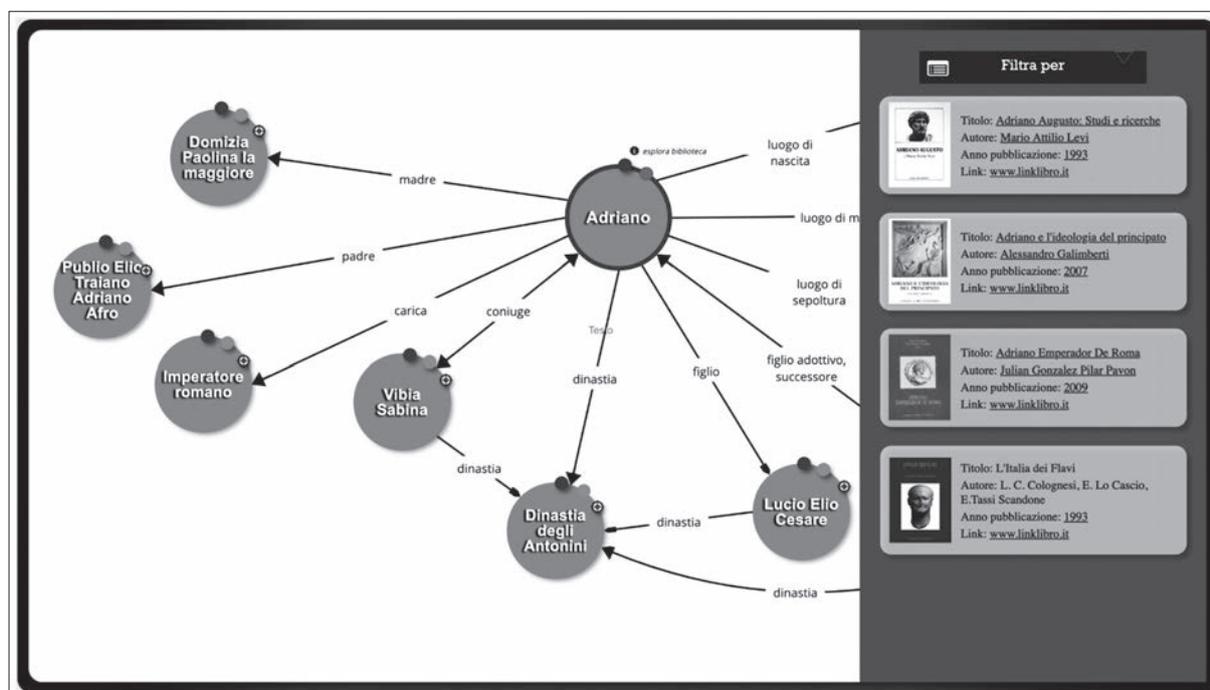


Fig. 2 – Mockup dell'interfaccia grafica (elab. CERIANI et al. 2020).

solo una libreria di testi e immagini di interesse, ma anche una vera e propria *storia*, una presentazione del tutto personalizzabile, sull'argomento oggetto di ricerca.

Il meccanismo di ricerca suggerito da ARCA intende amplificare l'esperienza dell'utente offrendo anche occasioni di approfondimento e di "scoperta" casuale di fonti, temi e materiali diversi, che possano arricchire di nuovi spunti le ricerche di base. Da qui deriva il potenziale effetto *serendipity* cui si è accennato (Fig. 3). Il principio dell'identificazione e dell'estrazione automatica delle entità semantiche dall'intero testo (non solo dai metadati di base) e la correlazione simultanea tra più testi, garantiti dal sistema grazie all'impiego di algoritmi di *AI*, rendono possibile non solo l'individuazione del maggior numero di risorse disponibili, ma anche la navigazione delle relazioni generate, progressivamente più lontane dal tema di ricerca originaria, ma tali da lasciare aperto il campo alla scoperta di connessioni inaspettate, utili per la ricerca in corso o comunque stimolanti per l'acquisizione di nuove conoscenze. Nei processi di *serendipity*, infatti, l'aspetto chiave risiede nella conoscenza accidentale, in questo caso generata da una rete di associazioni semantiche.

A bene vedere, l'effetto *serendipity*¹⁷ è in sé insito nella ricerca scientifica e come evidenziato dagli studi di sociologia e di psicologia cognitiva è fortemente connesso sia al desiderio di sviluppo e di innovazione sia al grado di avanzamento delle conoscenze. Per avviare il processo cognitivo di riconoscimento di dati anomali, imprevisti e strategici, occorrono quelli che Moretti¹⁸ definisce "occhi e menti allenate", particolarmente sollecitati quando operano in micro-ambienti socio-cognitivi stimolanti. Un aspetto interessante risiede nel valore assunto nell'ambito della divulgazione scientifica. Con le parole di Moretti¹⁹: "Il fatto che il concetto di *serendipity* sia nato al di fuori dei confini propri della ricerca sociologica e nel corso del tempo abbia raggiunto un pubblico sempre meno specialistico – con tutto ciò che questo significa sia dal versante della sua diffusione, di gran lunga più ampia di quella di norma riservata al sapere sociologico... – è, e resta, semplicemente un fatto. Esso non impedisce però di fare delle scelte, da variegati punti di vista ne sollecita anzi la necessità, e quelle qui operate mirano a focalizzare l'attenzione intorno alla reale, concreta, quotidiana attività di ricerca, fatta di rigore, di metodo, persino di ripetitività. Questioni di *inspiration and perspiration*, come avrebbe detto il vecchio Thomas Edison. Questioni, come lo stesso Merton racconta, di falsificazione scientifica in termini sociologici dello Standard Scientific Article (SSA), di differenze tra il modo personale di sviluppare i propri pensieri e l'ordine nel quale essi vengono presentati agli altri. In definitiva le discrepanze tra l'effettivo corso di un'indagine scientifica e la sua documentazione pubblica nel momento in cui il saggio o la monografia scientifica vengono presentati con quell'aspetto immacolato che poco o nulla lascia intravedere delle intuizioni, delle false partenze, degli errori, delle conclusioni approssimative e dei felici 'accidenti' che ingombrano il lavoro di ricerca, fanno sì che la documentazione pubblica della scienza non sia di per sé (in quanto tale) in grado di fornire parte significativa del materiale necessario alla ricostruzione del corso effettivo dello sviluppo scientifico".

Il sistema ARCA, dunque, non è inteso come un motore di ricerca bibliografica, bensì come un ambiente digitale dinamico in grado di offrire all'utente la possibilità di sperimentare molteplici percorsi per individuare testi e materiali di interesse. Procedimenti, percorsi, approcci, metodi e risultati possono poi essere annotati, conservati, raccontati ed eventualmente condivisi a seconda delle esigenze dell'utente, per il quale l'esperienza stessa di ricerca e di scoperta diventa una "storia".

¹⁷ Il termine entra nel vocabolario delle scienze umane nella seconda metà del Novecento grazie agli studi del filosofo Robert K. Merton. A seguito della sua approfondita analisi sociologica, è ormai condivisa la definizione di *serendipity*, non solo come semplice casualità della scoperta, ma accidentalità che lo scienziato deve saper cogliere e seguire, anche inconsapevolmente, e che conduce a scoperte fondamentali per l'avanzamento della conoscenza; MERTON 2002.

¹⁸ MORETTI 2016, pp. 5-10.

¹⁹ *Ibid.*

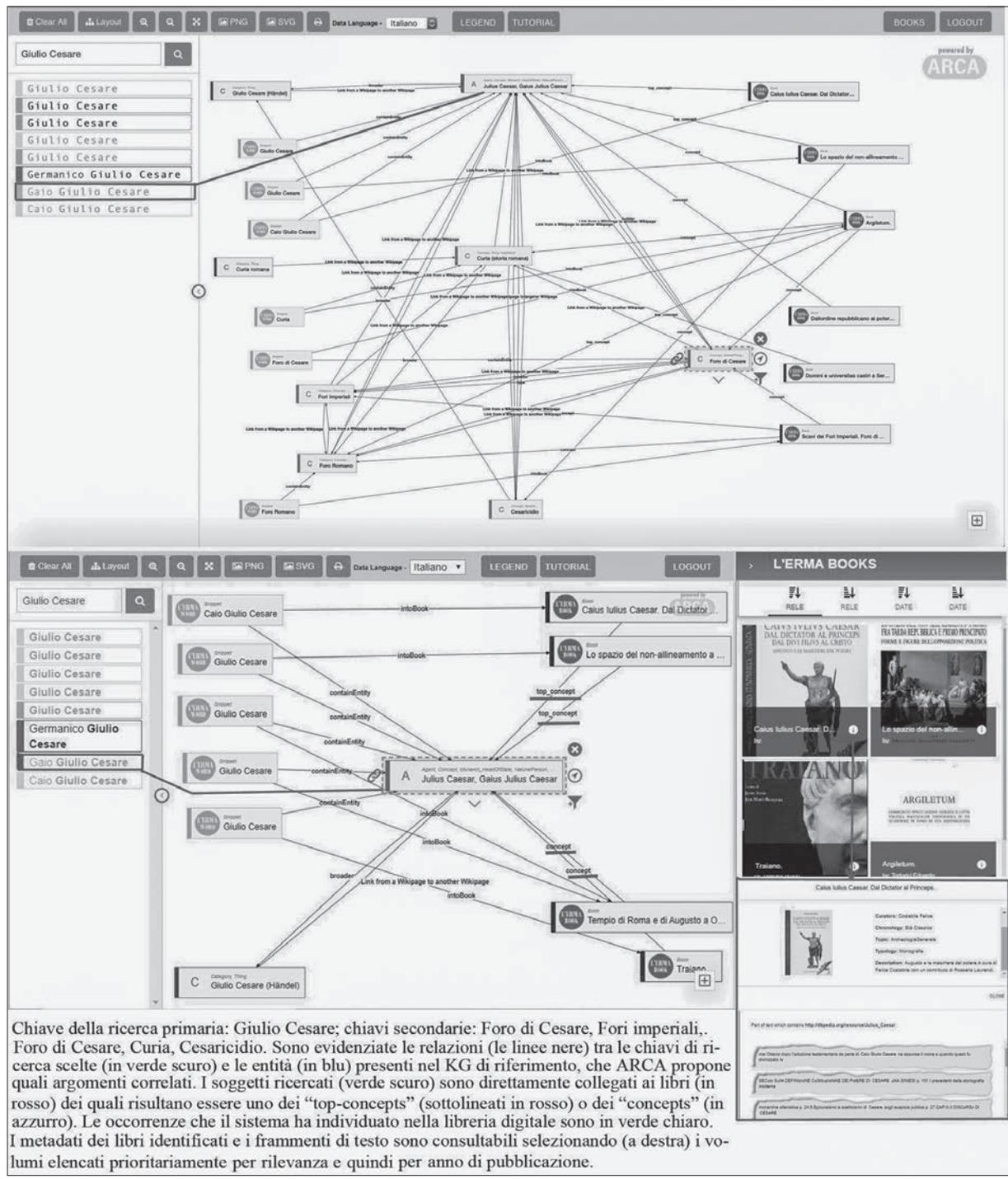


Fig. 3 – Un esempio di navigazione in ARCA (elab. M.C. Capanna).

L’approccio grafico-narrativo del *digital storytelling* per il suo carattere di immediatezza espressiva lascia un ampio margine alle relazioni concettuali; sul piano cognitivo i meccanismi di memorizzazione dell’utente si semplificano nei termini del rapporto di causa-effetto, ma si moltiplicano i nessi semantici. Il meccanismo di intertestualità, infatti, favorisce la reiterazione del principio per cui “una storia genera altre storie”, che a sua volta accelera la trasmissione delle conoscenze, il confronto, l’analisi critica e la genesi di ricerche inedite. In tal senso lo *storytelling* risponde all’esigenza non solo di condividere la propria esperienza, ma anche di divulgare acqui-

sizioni scientifiche, di trasmettere conoscenze specialistiche ed “educare alla scoperta”. Una storia digitale, come ogni storia, per essere una “buona storia” deve convincere, connettere e coinvolgere il lettore; se riesce nel suo intento essa facilita quella che si definisce *networked knowledge* (conoscenza connettiva) e sollecita la *combinatorial creativity* (creatività combinatoria).

Sia nel *digital* sia nel *visual storytelling*²⁰ resta fermo il presupposto di base della coerenza dello schema narrativo rispetto al valore e al significato del contenuto e in relazione alle aspettative e alle esigenze del target di riferimento. Il *visual storytelling*, di frequente impiegato anche come strumento didattico, sfrutta la capacità evocativa delle immagini per realizzare narrazioni ipermediali e consentire all’utente di stabilire collegamenti immediati, diretti o indiretti, tra nozioni e concetti. Sembrerebbe, infatti, che il cervello umano elabori le immagini visive 60.000 volte più velocemente e ricordi l’80% di ciò che vede, e solo il 20% di ciò che legge.

Nel *web marketing* culturale, sebbene in ritardo rispetto ad altri settori, sono in rapido sviluppo nuovi linguaggi digitali per la valorizzazione del patrimonio culturale e la divulgazione della conoscenza che sottendono un approccio integrato e trasversale nella comunicazione della cultura a diversi livelli di specializzazione. Uno strumento di grande utilità in tali processi comunicativi è il *digital storytelling*, impiegato anche nel sistema di ricerca semantica ARCA per affidare all’utente la creazione della propria “storia” di ricerca bibliografica, la gestione personalizzata e l’organizzazione creativa dei contenuti multimediali e dei risultati testuali di interesse, sui quali tornare ad interagire anche in fasi successive. Focalizzato su creatività e innovazione, il *digital storytelling* innesca meccanismi significativi nei processi tanto di acquisizione cognitiva quanto di rielaborazione collettiva della conoscenza. In questa medesima ottica le ricerche di tipo pedagogico, ad esempio, tendono a concentrarsi sull’analisi dei processi riflessività e dei cd. *lifelong learning*.

Nel sistema ARCA, progetto pilota rivolto in via sperimentale a studiosi ed esperti nel campo delle scienze antichistiche, le fonti dalle quali sono tratti i contenuti visivi, grafici e fotografici che vanno ad integrarsi ai risultati bibliografici e agli elementi testuali rispondono ai criteri di verificabilità, coerenza e attendibilità scientifica²¹. ARCA, dunque, conserva la continuità tra significato ed espressione, tra dato e comunicazione del dato. In merito alla scelta di forme di narrazione integrata, analoghe esperienze di carattere scientifico, finalizzate alla ricerca semantica per immagini tramite strumenti di comparazione, hanno restituito un quadro interessante rispetto alle potenzialità di ottimizzazione, sviluppo e innovazione di queste soluzioni tecnologiche che sfruttano applicazioni software *web-based*²².

C.d.F.

FORMALIZZAZIONE SEMANTICA DELLA CONOSCENZA E GRADO DI INNOVAZIONE IN ARCA

Secondo un’arguta osservazione di Tim Berners-Lee²³ “se l’interattività non significa soltanto stare seduti passivamente davanti a uno schermo, allora l’intercreatività non significa solo starsene seduti davanti a qualcosa di interattivo”. Tenendo a mente questo principio, si comprende quan-

²⁰ Il *web 2.0* offre numerose applicazioni e servizi che permettono la creazione di “storie digitali”. A seconda dei media e dei modelli con i quali si intende organizzare la narrazione di un tema o di una problematica si può scegliere una diversa tipologia di *storytelling*: *Timeline*; *Storymapping*; *Transmedia storytelling*; *Visual storytelling*; *Video storytelling*.

²¹ Le fonti, istituzionali e validate, da cui attingere i contenuti multimediali e le immagini da impiegare nello *storytelling* si rintracciano nelle piattaforme *open-source* e *open-access* e nei maggiori *repository* nazionali e internazionali. In generale sull’argomento, e sull’uso di *Linked open Data*, per la diffusione dei prodotti della ricerca scientifica, SOLODOVNIK 2015.

²² Vd., ad esempio, <https://iiif.io>.

²³ BERNERS-LEE 1999.

to la proposta di ARCA di fornire un sistema integrato sia orientata verso modalità di fruizione dinamica e creativa dei contenuti, che superino la staticità delle edizioni digitali trasformate in documenti associabili, grazie alle potenzialità del *web* semantico e all'uso dei *Linked Data*, a una più ampia rete di informazioni condivise e interconnesse.

Il processo di formalizzazione semantica della conoscenza, mediante interpretazione e rielaborazione automatica dei dati associati in insiemi semantici, genera risultati di ricerca più raffinati e consente la creazione di reti connettive tra documenti e testi. Inoltre, l'estrazione di entità dai *KG* restituisce ontologie di metadati potenzialmente in grado di ampliare il campo di ricerca, eliminando il principio di casualità nei risultati. L'ideazione di un sistema implementabile sulla base dell'esperienza degli utenti e il rovesciamento di prospettiva da una fruizione individuale a una fruizione condivisa e integrata offrono al ricercatore il vantaggio di poter confrontare i propri risultati con percorsi bibliografici già positivamente sperimentati e verificati dai colleghi. La *user experience* è inoltre favorita dalla adattabilità degli strumenti di esplorazione alle esigenze di ricerca non solo per parola chiave, ma anche per concetto, data l'opportunità di interrogare la *RoadMap* ed esplorare, contemporaneamente, la biblioteca virtuale e il grafo. I contenuti di interesse rintracciati sono poi organizzabili secondo le esigenze dell'utente/ricercatore.

In conclusione, gli elementi che contribuiscono all'efficacia di un sistema integrato a un motore di ricerca semantica, e i vantaggi che questo potrebbe apportare negli studi accademici sono:

- Uso di *KG* e Ontologie di metadati istituzionali
- Garanzia di scientificità delle fonti sorgente
- Riduzione del *time-to-search*
- Ottimizzazione dei risultati
- Agevolazione nella *user experience*
- Associazioni semantiche e relazioni tra dati utili per confronti bibliografici o tematici
- Articolazione e integrazione dei contenuti per successivi livelli di approfondimento
- Fruizione dinamica, creazione di note, elenchi, bibliografie tematiche
- Conservazione e integrazione dei risultati delle ricerche
- Condivisione e divulgazione dei prodotti delle ricerche scientifiche ad ampio raggio
- Reperimento diretto delle informazioni editoriali e commerciali

Rispetto alle soluzioni software esistenti, il prototipo ARCA, implementabile e aperto a futuri sviluppi, impiega un *KG* aperto ed enciclopedico, applicandolo a uno specifico ambito disciplinare, e un'interfaccia composita, che include ricerca testuale, esplorazione visuale del grafo, visualizzazione articolata delle entità e libreria di testi. L'idea di convertire il prodotto digitale in una risorsa continuamente interrogabile in chiave interdisciplinare, ferma restando la verifica della scientificità delle fonti sorgente, rappresenta una innovazione finora sconosciuta nel settore umanistico e offre l'occasione di aderire alle attuali forme di comunicazione digitale e ai più moderni sistemi di collegamento e condivisione dei dati su scala globale.

M.C.C., C.d.F.

Maria Cristina Capanna
mariacristina.capanna@uniroma1.it

Clara di Fazio
clara.difazio@uniroma1.it

Abbreviazioni bibliografiche

BERARDI *et al.* 2003: D. BERARDI - D. CALVANESE - G. DE GIACOMO - M. LENZERINI - M. MECELLA, *Automatic Composition of E-services that Export their Behavior*, in M.E. ORLOWSKA - S. WEERAWARANA - M.P. PAPAZOGLU - J. YANG (eds.), *Service-Oriented Computing - ICSOC 2003, Lecture Notes in Computer Science*, vol. 2910, Springer-Berlin-Heidelberg 2003.

BERNERS-LEE 1999: T. BERNERS-LEE, *Weaving the Web. The Original Design and Ultimate Destiny of the World Wide Web*, 1999.

BIAGETTI 2019: M.T. BIAGETTI, *Le biblioteche digitali: tipologie, funzionalità e modelli di sviluppo*, Milano 2019.

BIANCHI *et al.* 2017: F. BIANCHI - M. PALMONARI - M. CREMASCHI - E. FERSINI, *Actively Learning to Rank Semantic Associations for Personalized Contextual Exploration of Knowledge Graphs*, in Proceedings, Part I 14th International Conference, ESWC 2017 (Portorož-Slovenia 2017), pp. 120-135.

BIRD *et al.* 2009: S. BIRD - E. KLEIN - E. LOPER, *Natural Language Processing with Python*, O'Reilly Media 2009.

BOD 2013: R. BOD, *A New History of the Humanities. The Search for Principles and Patterns from Antiquity to the Present*, Oxford 2013.

BOD 2019: R. BOD, *Le scienze dimenticate. Come le discipline umanistiche hanno cambiato il mondo*, Roma 2019.

CERIANI *et al.* 2020: M. CERIANI - E. BERNASCONI - M. MECELLA, *A Streamlined Pipeline to Enable the Semantic Exploration of a Bookstore*, in M. CECI - S. FERILLI - A. POGGI (eds.), *Digital Libraries: The Era of Big Data and Data Science-IRCDL 2020 Communications in Computer and Information Science*, vol. 1177, Springer-Cham 2020, pp. 75-81.

MAYNARD *et al.* 2017: D. MAYNARD - K. BONTICHEVA - I. AUGENSTEIN, *Natural Language Processing for the Semantic Web*, Morgan Claypool 2017.

MARRELLA *et al.* 2017: A. MARRELLA - M. MECELLA - S. SARDINA, *Intelligent process adaptation in the SmartPM system*, in *ACM Transactions on Intelligent Systems and Technology (TIST)* 8.2, 2017, pp. 1-43.

MERTON 2002: R.K. MERTON (con E.G. Barber, trad. it. M.L. Bassi), *Viaggi e avventure della serendipity: saggio di semantica sociologica e sociologia della scienza*, Bologna 2002.

MORETTI 2016: V. MORETTI, *Intorno al concetto di Serendipity e alle sue prospettive nell'ambito della ricerca scientifica*, in *Scienze e Ricerche* 33, 2016, pp. 5-10.

PALMONARI *et al.* 2015: M. PALMONARI - G. UBOLDI - M. CREMASCHI - D. CIMINIERI - F. BIANCHI, *DaCENA: Serendipitous News Reading with Data Contexts*, in *Revised Selected Papers - The Semantic Web*, ESWC 2015 Satellite Events (Portoroz-Slovenia 2015), pp. 133-137.

SOLODOVNIK 2015: I. SOLODOVNIK, *Repository Istituzionali Open Access e strategie Linked Open Data. Per una migliore comunicazione dei prodotti della ricerca scientifica*, Firenze 2015.

ABSTRACT

ARCA Project (*Academic Research Creativity Archives*) offers an innovative digital solution for advanced semantic and bibliographic research. Focused on interdisciplinarity, integration and sharing of technical-scientific knowledge and skills between *technologies* and *humanities*, it suggests the combination of research methods of humanities with the tools of the ICT world (*Information and Communication Technologies*). The aim of ARCA is to create an integrated platform of editorial products and digital services that can offer new semantic and bibliographic research methods and experiences, renewing the traditional method.

Oriented to current forms of development and multimedia production, cross-media distribution and digital post-production, ARCA platform makes possible a creative use of academic contents, transformed into associable documents, overcoming the static nature of digital editions. Thanks to the semantic web ARCA creates a wider network of shared and interconnected information.

Edizioni Quasar di Severino Tognon s.r.l.
via Ajaccio 41/43 – 00198 Roma
tel. 0685358444, fax 0685833591
www.edizioniquasar.it

per informazioni e ordini
qn@edizioniquasar.it

ISSN 1123-5713

ISBN 978-88-5491-078-2

Finito di stampare nel mese di settembre 2020
presso Global Print – Gorgonzola (MI)