

AI, data management e Web mining per l'arte e la cultura

Marco Schaerf



SAPIENZA
UNIVERSITÀ DI ROMA

Tecnologie e studi umanistici

- Area in grande sviluppo ma ancora poco strutturata
- Mancano figure esperte di entrambi i campi
- Diversità di linguaggio, metodologie di analisi e stili di pubblicazione

Iniziative in Sapienza e Lazio

- Sapienza da anni ha costituito centro Digilab nel 2012 dove afferiscono ricercatori dei dipartimenti umanistici e tecnologici
- Regione Lazio ha avviato il Distretto delle tecnologie per la cultura DTC dal 2018

Intelligenza artificiale

- Insieme di metodi e tecniche per rendere «intelligente» l'elaborazione dell'informazione
 - Approcci simbolici: rappresentazione della conoscenza, ontologie, ragionamento automatico
 - Approcci black-box: machine learning, Deep Neural Networks
 - Una nuova categoria ? IA generativa

Web mining

- Insieme di tecniche ed approcci per raccogliere e sistematizzare le informazioni open sul web, con lo scopo di creare dei data lake su cui inferire proprietà interessanti
 - *Social signal processing* : comprendere i mood a partire dalle pubblicazioni sui social

Data management

- Insieme di tecniche per gestione efficace ed efficiente delle informazioni, non solo strutturate ma e soprattutto multimediali
 - Datalake
 - Datalake house
 -

SM&ST (1)

- Smart Museum & Smart Tourism
 - Progetto »cluster» del 2012
 - Vicenda travagliata
- Innovazioni introdotte
 - Architettura innovativa di data management basata su dati strutturati e non (*datalake* prima che si usasse questo termine) *on top* della quale utilizzare ontologie per la gestione dell'informazione

SM&ST (2)

- Architettura a micro-servizi, anche asincroni, per lo scambio efficace di informazioni
- Approccio UCD - User Centered Design per la produzione di Web app utilizzabili dal turista medio
- Idee iniziali di quello che poi si vedrà in altri progetti

Digital Library Sapienza

- Progetto di raccogliere e catalogare tutto il materiale bibliografico ed archivistico di Sapienza con tutte le informazioni collegate
- Sviluppo di una classificazione compatibile con i più avanzati schemi di classificazione

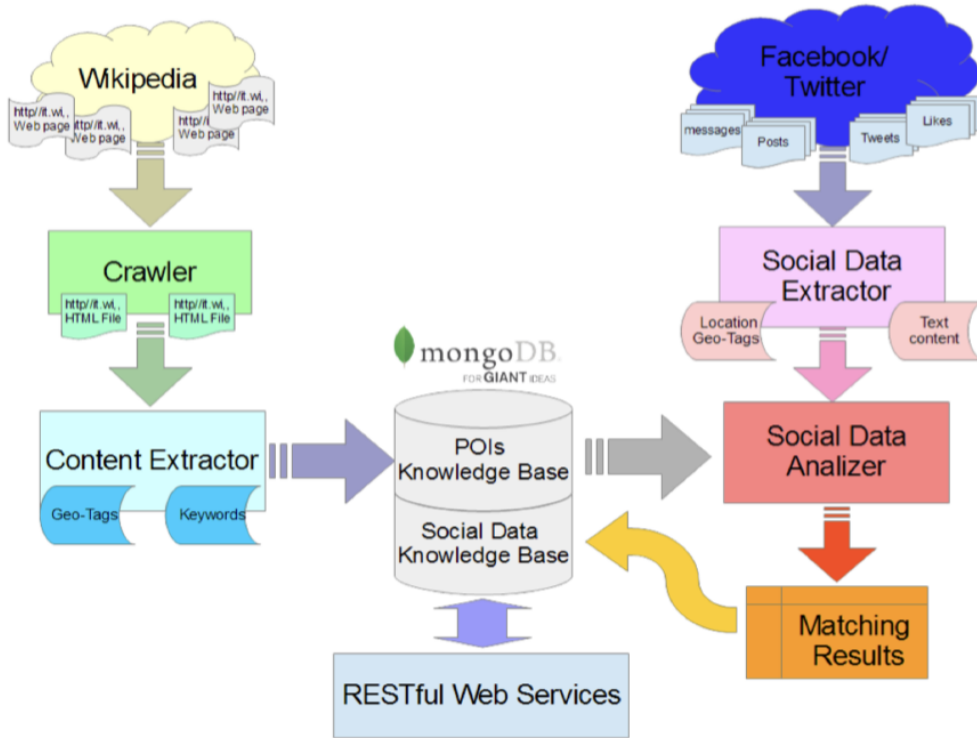
Digital Library Sapienza: Problemi

- Complessità della digitalizzazione e della classificazione
- Costo della digitalizzazione
- Costo dell'infrastruttura
- «Gelosia» dei propri dati

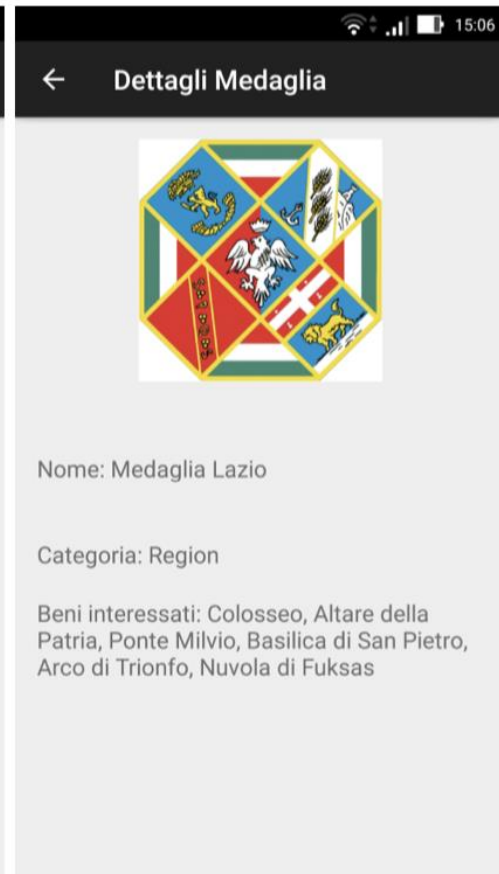
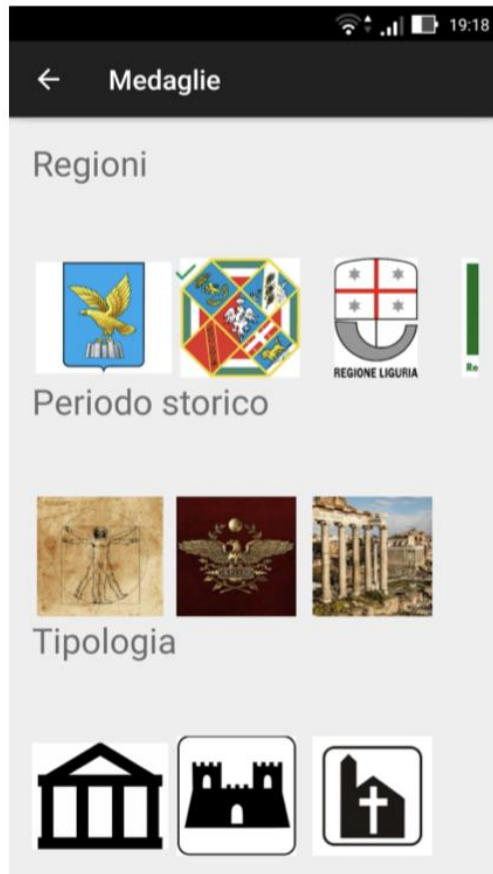
NEPTIS (1)

- Progetto PON 2015 - 2016
- Innovazioni introdotte
 - Uso del social signal processing / web mining per suggerire cosa vedere di un sito, utile soprattutto per i siti minori
 - Uso di tecniche di gamification per incentivare la visita a particolari categorie di utenti

NEPTIS (2)



NEPTIS (3)



CiceroNet / SivalaNet

- Progetti della fase 2 del Centro di Eccellenza Distretto Beni Culturali Regione Lazio
- Suggerimenti di percorsi ottimizzati (attraverso tecniche di planning) che tengono conto di dati raccolti sul Web
 - Adatti ai beni/siti minori ed al cosiddetto business traveller
 - CiceroNet: focus sui percorsi ciceroniani nel Lazio
 - SivalaNet: focus sull'Etruria laziale

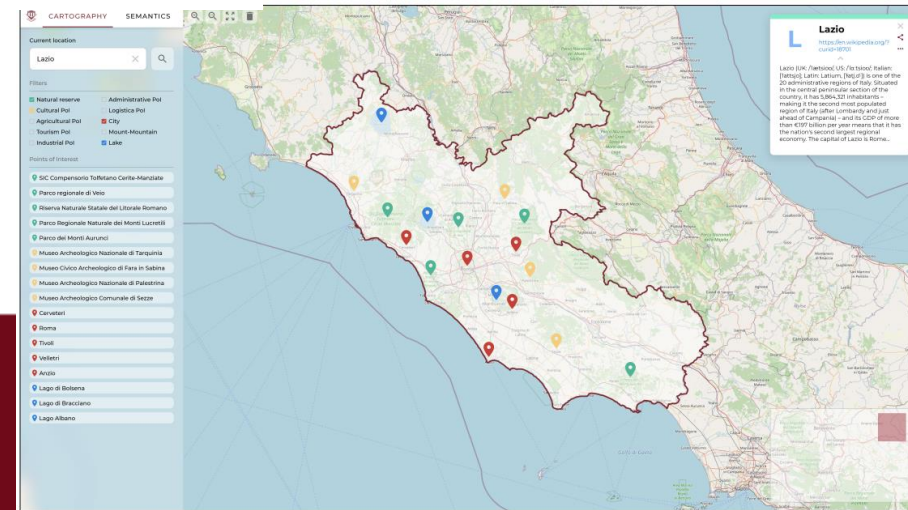
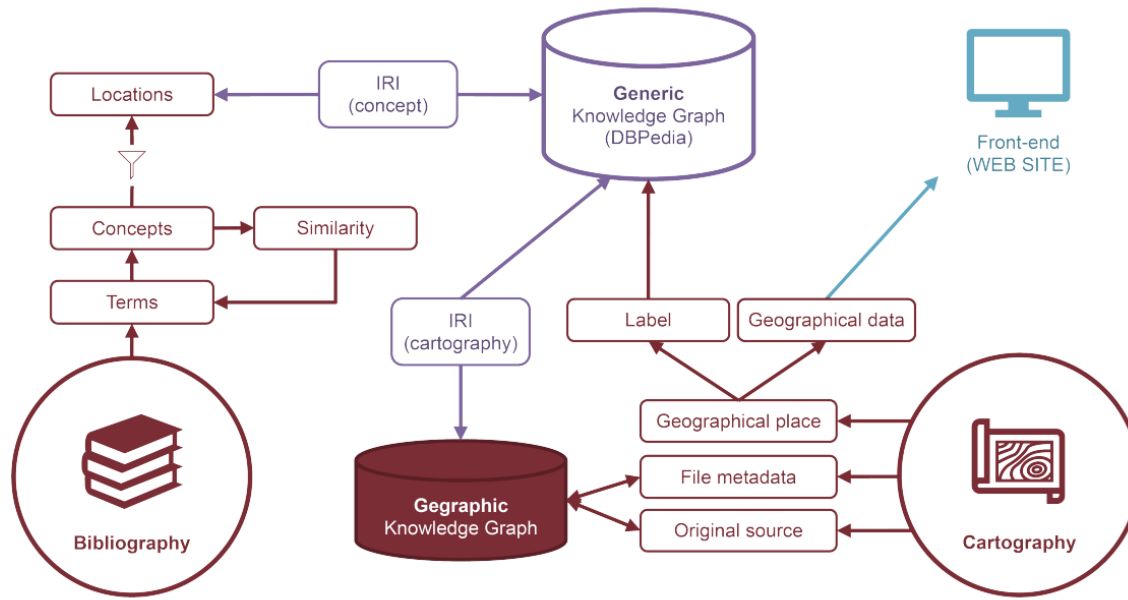
ARCA/SCIBA (1)

- Search engine specifico per studiosi archeologici
- Innovazione: knowledge graph costruiti sulla base di dati estratti da documenti e post-georeferenziati (nel caso di SCIBA)

ARCA/SCIBA (2)

(1)
(2)

ARCA/SCIBA (3)



Artemisia

- Studio dei flussi e dei pattern di visita dei musei
- Analisi basata su sensoristica non invasiva (no video) e comportamenti sociali
- Difficoltà acquisizione grandi moli di dati

Criticità incontrate

- Il tipico studioso di humanities ha molta diffidenza verso tecniche automatiche di gestione dell'informazione (e.g. DNN) che non garantiscono il 100% di scientificità
 - Attenzione all'utente del sistema, se un *knowledgeable* e quindi a scopo divulgativo od un *expert* a scopo di supporto alla propria ricerca

Nuove figure professionali

- Creare esperti di digital humanities con il corretto approccio
- Vari corsi di laurea in Italia
- In Sapienza Filosofia ed Intelligenza Artificiale
- Manca chiaro inquadramento (problema SSD, aree, comunità accademica ...)