



UNIVERSITÀ DEL PIEMONTE ORIENTALE

Linked open data non basta:
computer, intelligenza,
connessioni tra i dati

Maurizio Lana

open data

- il tema nasce nel contesto pubblico politico/amministrativo per la cittadinanza attiva (conoscere, controllare, agire)
- si amplia alla ricerca: Open Access non solo per le pubblicazioni ma anche per i dati su cui esse si basano
 - richiesto dai bandi europei
 - contemplato dalle licenze CC

i link del web

- due tipi di link ipertestuali del web:
 - connessioni generate automaticamente: collegano dati connessi a priori
 - cfr. i link dentro una libreria online
 - connessioni costruite per effetto di intenzionalità esplicita: collegano “dati” oggetto di analisi, valutazione e interpretazione

link intenzionali

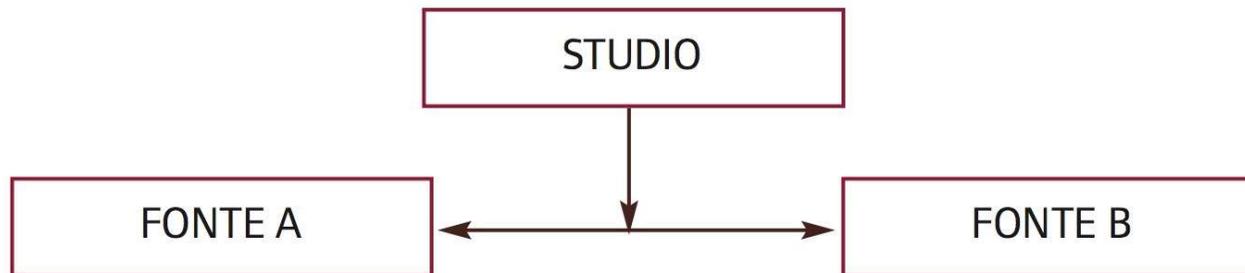
- ipertesti, scritti che sfruttano le possibilità argomentative offerte dal mondo digitale
- collegando due differenti fonti si compie un atto interpretativo: due entità distinte diventano parte di un'unità più ampia

un caso modello di interpretazione

- 2 fatti storici:
 - assassinio di Ahmad Shāh Massoūd in Afghanistan il 9 settembre 2001
 - attacco alle torri gemelle del World Trade Center l'11 settembre 2001
- probabilmente collegati da una comune matrice qaidista
- come si esprime nel web l'argomentazione su questo tema?

interpretazione nel web

- il link tra pagine che citano il primo e il secondo evento non basta perché non ci sono typed links nel web
- occorre un atto interpretativo esplicito che crea e motiva la connessione



approfondimento: typed links nel web

- il web sarà “a large hypertext database with typed links”, scriveva Berners Lee nel 1989 in *Information management. A proposal*,
<<http://www.w3.org/History/1989/proposal.html>>
- successiva discussione nel documento senza data (ma forse è del 1990) <<http://www.w3.org/DesignIssues/Topology.html>>

interpretazione di Massoud e 11/9/11

- due esempi tra tanti:
- https://en.wikipedia.org/wiki/Ahmad_Shah_Massoud
- <http://www.aljazeera.com/indepth/opinion/2014/09/afghanis-tan-shadow-ahmad-shah-mas-2014997826874331.html>
- l'interpretazione nel web generalmente usa il link in modo confermativo
 - l'affermazione è linkata alla fonte che la giustifica o dettagliandola o approfondendola

interpretazione di Massoud e 11/9/11

- la connessione LOD tra gli eventi citati può avvenire automaticamente per la presenza dell'espressione al qaeda
- ma la conoscenza si esprime nel tradizionale link ipertestuale tra tali eventi, che esplicita un atto di interpretazione storica operato da un'intelligenza umana

ambiti dei linked open data (1)

- gli open data della spesa sanitaria di una Regione per essere compresi richiedono dati di bilancio della Regione
 - per entrambi i tipi di dati la comprensibilità (la interpretabilità) migliora
 - se se ne possiede una serie storica
 - se sono disponibili altri dati aperti del medesimo tipo di altre Regioni
- la connessione tra i dati si basa sulla conoscenza del dominio posseduta da coloro che di questi dati si interessano

ambiti dei linked open data (2)

Hackaton 2015 @ NEXA9

- dati.piemonte.it contiene open data sui musei del Piemonte, con informazioni testuali semi-strutturate nel campo «orari di apertura»
- un'applicazione estrae e ordina i dati sulle chiusure festive, le aperture straordinarie e serali, il tipo di ingresso (a pagamento/gratuito)
- a quel punto si generano i LOD: si collegano i musei in base al tipo e orario di apertura (ordinaria, festiva, serale), ingresso, e così via
 - un foglio Excel ben progettato potrebbe bastare, ma le info si trovano in fonti distinte e indipendenti
- il procedimento di creazione del link è indiscutibile ed ovvio: si basa sul riordino di informazione già esistente

quantità vs. qualità

- STMicroelectronics e la Cattedrale di S. Agata sono collegate attraverso Geonames in quanto situate in Catania ma ciò è di per sé irrilevante perché migliaia di entità sono interconnesse tramite Catania
- diventerebbe rilevante se portasse all'evidenza una notizia su STMicroelectronics che finanzia restauri della Cattedrale perché intende essere una presenza culturale sul territorio
effetto "finale dei predatori dell'arca perduta"...

intersezione tra interpretazione umana e LOD

- il progetto Geolat in corso all'UPO integra LOD e interpretazione attraverso la creazione di un'ontologia geografica che permette di annotare semanticamente nomi geografici
- l'annotazione contiene link LOD ad un gazzettiere ma grazie all'ontologia è possibile esprimere conoscenza riguardante il luogo

in conclusione

produrre LOD con procedure automatiche sostanzialmente incontrollate non solo non elimina, ma sottolinea la necessità di selezionare dati che rispondano ad un'ipotesi interpretativa, per poi connetterli e pubblicarli affinché l'ipotesi comunicata e condivisa possa essere verificata, convalidata, riusata in altri contesti di ricerca