

Sistemi basati su ontologie per le Digital Humanities

Carlo Meghini, Valentina Bartalesi, Filippo Benedetti, Daniele Metilli

CNR ISTI

{first.last} at isti.cnr.it

1 Introduzione

Un sistema informativo per le Digital Humanities (DH per brevità) è visto come un agente che manipola conoscenza, essendo in grado di accogliere conoscenza dal mondo esterno (funzione di TELL) e di elaborarla in modo appropriato per rispondere a domande relative alla conoscenza posseduta (funzione di ASK). Questa visione non è originale: essa è stata proposta all'inizio degli anni 80 come paradigma di sviluppo di agenti intelligenti, in una serie di articoli scientifici seminali, a partire dall'indirizzo dell'allora Presidente dell'Associazione Nazionale Americana dell'Intelligenza Artificiale, Prof. Allen Newell [Newell, 1982]. A questo articolo ha fatto seguito una imponente serie di articoli dei maggiori studiosi (si veda per esempio [Levesque and Lakemeyer, 2001]). La visione pone la logica matematica al centro del sistema informativo, e in particolare la relazione di implicazione logica come quella che meglio cattura il rapporto tra domanda e risposta. Facendo ciò, Newell fa propria un'intuizione del rapporto tra logica e razionalità che è alla base dell'Intelligenza Artificiale sin dagli inizi.

La visione di Newell ha dato molti frutti. Oltre alla vasta letteratura sulle logiche per l'Intelligenza Artificiale, basterà ricordare che uno degli standard del web semantico [Berners-Lee *et al.*, 2001], il Web Ontology Language (OWL) [Motik *et al.*, 2009], è basato sui risultati ottenuti nel settore delle Logiche Descrittive da ricercatori che hanno cercato di realizzare la visione investigando logiche per la rappresentazione della conoscenza che fossero decidibili e possibilmente trattabili [Baader *et al.*, 2003].

Il nostro gruppo ha fatto propria questa visione e sta lavorando da qualche anno ad applicarla nel settore delle DH. La Sezione successiva descrive due progetti in cui tale visione si è esplicitata. Nel resto della sezione presente, si cercherà di argomentare perché la visione di Newell risulti particolarmente fruttuosa per le DH, al di là dei vantaggi che essa porta da un punto di vista ingegneristico.

Un tratto fondamentale delle DH è il loro essere al crocevia delle discipline umanistiche da un lato e dell'ingegneria dei sistemi informatici e della matematica computazionale dall'altro. Tradizionalmente, queste due branche sono considerate antitetiche, e così di fatto sono trattate nella grande maggioranza dei *curricula*. Questa drastica separazione rende difficile il cammino delle DH, per l'oggettiva difficoltà di comunicazione che essa crea tra chi dovrebbe necessariamente

collaborare allo sviluppo di questo settore delle scienze, proponendo tecniche e metodiche innovative e testando tali tecniche e metodiche in sistemi sperimentali. In particolare, riesce difficile trovare il terreno in cui la collaborazione tra umanisti e informatici debba attuarsi: le reciproche invasioni di campo portano a risultati discutibili come la scarsa usabilità dei sistemi disegnati dagli informatici o l'inappropriatezza di strumenti informatici impiegati da parte degli umanisti, quale l'uso dell'XML per la formalizzazione del testo.

In questo contesto, la scelta di intendere un sistema informativo come un agente in grado di manipolare conoscenza porta come naturale conseguenza una chiara separazione dei ruoli nella progettazione di tali sistemi: agli umanisti è dato il compito di stabilire che tipo di conoscenza debba manipolare il sistema, e quali ragionamenti debbano essere effettuati su tale conoscenza; agli informatici è dato il compito di tradurre tali direttive in algoritmi e strutture dati, impiegando le tecnologie di volta in volta più adeguate ai parametri di contesto. Terreno di incontro tra le due figure sono le ontologie, dove i concetti del dominio sono definiti congiuntamente, usando il linguaggio comune della logica e in modo del tutto indipendente dalle tecnologie. Il processo di design che viene così a stabilirsi è autenticamente cooperativo: il sistema risultante non nasce da qualcuno che ordina e qualcun altro che esegue: la stessa visione del sistema basato sulla conoscenza deve spesso essere trasmessa agli umanisti con dimostrazioni che semplifichino il concetto; come pure è desiderata degli umanisti sono meglio resi espliciti agli informatici attraverso il ricorso ad altri sistemi. Il dialogo procede per fasi successive fino al raggiungimento di un accordo formulato in termini delle entità di dominio e delle operazioni che su di esse sono familiari degli umanisti, indipendentemente dalla tecnologia con cui andranno a essere realizzate.

I due progetti illustrati nella prossima sezione sono nati da questa metodologia di lavoro.

2 Progetti

2.1 DanteSources e le fonti dantesche

DanteSources è un'applicazione web che consente l'accesso libero alla conoscenza delle fonti primarie di Dante Alighieri, cioè le opere di altri autori che Dante cita nei suoi testi. L'applicazione web è stata sviluppata all'interno del progetto PRIN 2013-2016 Per un'enciclopedia dantesca digitale, che

mira a costruire un prototipo di biblioteca digitale dotata di servizi di supporto agli studiosi per creare, arricchire e consultare un'enciclopedia digitale delle opere di Dante Alighieri. Fino ad oggi, la conoscenza delle fonti primarie di Dante Alighieri è stata raccolta in libri tradizionali e ciò ha reso difficile per gli studiosi recuperarla e produrne una panoramica completa. Usando le tecnologie del web semantico abbiamo sviluppato un'ontologia [Bartalesi and Meghini, 2017] per la rappresentazione semantica della struttura delle opere di Dante e della conoscenza relativa alle loro fonti primarie. L'ontologia è espressa in Resource Description Framework (RDF), un vocabolario raccomandato dal World Wide Web Consortium [Manola *et al.*, 2004] che fornisce i termini per rappresentare la conoscenza in formato *machine-readable*. Abbiamo quindi sviluppato un tool semi-automatico per supportare gli studiosi nel popolare l'ontologia con la conoscenza inclusa nei commenti alle opere di Dante scritti da autorevoli studiosi della materia. Questo tool salva automaticamente i dati inseriti in un *triple store* [Rohloff *et al.*, 2007] sotto forma di grafo RDF. Sulla base dell'ontologia, abbiamo sviluppato DanteSources, un'applicazione web che consente agli utenti di estrarre e visualizzare le informazioni memorizzate nel grafo come grafici e tabelle. L'ontologia e gli strumenti che abbiamo sviluppato sono facilmente riutilizzabili, ad esempio per rappresentare la conoscenza relativa alle fonti primarie di altri autori della letteratura italiana ed internazionale.

L'ontologia e l'applicazione web sono disponibili al seguente indirizzo: <http://dantesources.org>

2.2 Narrazioni e biblioteche digitali

Le attuali biblioteche digitali (BD) sono per lo più costruite intorno a grandi collezioni di oggetti scarsamente correlati. La nostra ricerca mira ad arricchire lo spazio informativo delle BD introducendo le narrazioni, entità intese come reti di eventi collegati tra loro attraverso relazioni semantiche. Ciascun evento, inoltre, è collocato nello spazio e nel tempo e collegato ai propri componenti fattuali (es. persone, oggetti ecc.). Per introdurre le narrazioni così intese in una BD, abbiamo sviluppato un'ontologia [Bartalesi *et al.*, 2017] basata sui concetti fondamentali della scienza narrativa. L'ontologia è espressa usando come vocabolario di riferimento il CIDOC CRM [Doerr, 2003], e codificata in OWL 2 DL. In particolare, l'ontologia permette di: (i) rappresentare gli eventi che compongono una narrazione, collegandoli attraverso tre tipi di relazioni semantiche: di occorrenza temporale, di causa e mereologica; (ii) collegare un evento con i relativi oggetti digitali inclusi in una BD; (iii) rappresentare la *provenance* dei dati per le narrazioni storiche. La versione corrente dell'ontologia non è definitiva: stiamo infatti lavorando per rappresentare anche il livello testuale della narrazione. Sulla base dell'ontologia abbiamo sviluppato un tool semi-automatico per la costruzione e la visualizzazione di narrazioni [Metilli *et al.*, 2019]. Il tool recupera le istanze delle classi dell'ontologia usando Wikidata [Vrandečić and Krötzsch, 2014] come una base di conoscenza esterna e facilita la costruzione e la contestualizzazione degli eventi e il loro collegamento. La conoscenza collezionata viene poi automaticamente salvata come grafo OWL in un *triple store*. Quindi il tool consente di visualizzare le conoscenze contenu-

ta nel grafo in formati facilmente accessibili dall'utente come tabelle e timeline.

L'ontologia e il tool sono disponibili al seguente indirizzo: <https://dlnarratives.eu/>

Riferimenti bibliografici

- [Baader *et al.*, 2003] Franz Baader, Diego Calvanese, Deborah L. McGuinness, Daniele Nardi, and Peter F. Patel-Schneider, editors. *The Description Logic Handbook: Theory, Implementation, and Applications*. Cambridge University Press, 2nd edition, 2003.
- [Bartalesi and Meghini, 2017] Valentina Bartalesi and Carlo Meghini. Using an ontology for representing the knowledge on literary texts: The dante alighieri case study. *Semantic Web*, 8(3):385–394, 2017.
- [Bartalesi *et al.*, 2017] Valentina Bartalesi, Carlo Meghini, and Daniele Metilli. A conceptualisation of narratives and its expression in the crm. *International Journal of Metadata, Semantics and Ontologies*, 12(1):35–46, 2017.
- [Berners-Lee *et al.*, 2001] Tim Berners-Lee, James Hendler, Ora Lassila, et al. The semantic web. *Scientific american*, 284(5):28–37, 2001.
- [Doerr, 2003] Martin Doerr. The cidoc conceptual reference module: an ontological approach to semantic interoperability of metadata. *AI magazine*, 24(3):75, 2003.
- [Levesque and Lakemeyer, 2001] Hector J. Levesque and Gerhard Lakemeyer. *The Logic of Knowledge Bases*. The MIT Press, 2001.
- [Manola *et al.*, 2004] Frank Manola, Eric Miller, Brian McBride, et al. Rdf primer. *W3C recommendation*, 10(1-107):6, 2004.
- [Metilli *et al.*, 2019] Daniele Metilli, Valentina Bartalesi, and Carlo Meghini. A wikidata-based tool for building and visualising narratives. *International Journal on Digital Libraries*, pages 1–16, 2019.
- [Motik *et al.*, 2009] Boris Motik, Peter F Patel-Schneider, Bijan Parsia, Conrad Bock, Achille Fokoue, Peter Haase, Rinke Hoekstra, Ian Horrocks, Alan Ruttenberg, Uli Sattler, et al. Owl 2 web ontology language: Structural specification and functional-style syntax. *W3C recommendation*, 27(65):159, 2009.
- [Newell, 1982] Allen Newell. The knowledge level. *Artificial Intelligence*, 18(1):87–127, 1982.
- [Rohloff *et al.*, 2007] Kurt Rohloff, Mike Dean, Ian Emons, Dorene Ryder, and John Sumner. An evaluation of triple-store technologies for large data stores. In *OTM Confederated International Conferences On the Move to Meaningful Internet Systems*, pages 1105–1114. Springer, 2007.
- [Vrandečić and Krötzsch, 2014] Denny Vrandečić and Markus Krötzsch. Wikidata: a free collaborative knowledge base. 2014.