

Strutturazione e condivisione della conoscenza, informatica ed economia della catalogazione

Giovanni Antonio Cignoni

Enrico Meloni

Progetto HMR, Corso di Storia dell'Informatica, c/o Dipartimento di Informatica, Università di Pisa, Largo B. Pontecorvo, 3, I-56127 Pisa. E-mail: giovanni.cignoni@di.unipi.it; enrico.meloni@outlook.it

RIASSUNTO

Fra i tanti vantaggi della rivoluzione digitale ci sono anche il trattamento strutturato delle informazioni e gli strumenti di collaborazione per condividere conoscenza e competenze. Se ci restringiamo a un settore particolare del patrimonio culturale, sono oggi disponibili tecnologie sia per riunire in un'unica base di conoscenza le informazioni di catalogazione e documentazione di tutte le collezioni afferenti a quel settore, sia per gestire tale conoscenza in collaborazione con tutti gli esperti di quel dominio. Rispetto ai tradizionali cataloghi-elenchi (un reperto, una scheda), una base di conoscenza condivisa offre vantaggi cruciali: esplicita le relazioni, elimina replicazione e inconsistenza, abilita il "distant reading", cioè l'osservazione delle informazioni nel loro insieme. Un tale strumento permette a conservatori e curatori di concentrare il loro impegno sulle collezioni: le informazioni nella base di conoscenza sono di riferimento per identificare i reperti, per condurre ricerche storiche, per costruire la documentazione di un'esposizione. La condivisione delle informazioni è garanzia di affidabilità dei contenuti e mezzo di collaborazione fra colleghi. Computing History Knowledge Base (CHKB) è un sistema software per riunire in un'unica base di conoscenza le informazioni di tutte le collezioni afferenti a un settore particolare del patrimonio scientifico: la storia dell'informatica. CHKB realizza una doppia economia: è un uso pieno delle tecnologie disponibili e permette a curatori e conservatori di sfruttare al meglio il loro tempo. Il modello di CHKB è applicabile ad altri domini del patrimonio culturale.

Parole chiave:

base di conoscenza, patrimonio tecnico-scientifico, lavoro cooperativo, storia dell'informatica.

ABSTRACT

Structuring and sharing of knowledge, information technologies and economy of cataloguing

Structured managing of information and sharing of expertise through cooperative work are two of the many benefits resulting from the digital revolution. If we focus on a particular field of cultural heritage, viable technologies exist to set up a single knowledge base of all cataloguing and documentation information about all collections related to the given field. Moreover, it is possible to engage all the experts of the domain in the cooperative management of such knowledge. With respect to the traditional way of cataloguing (one record for each piece), a shared knowledge base has valuable benefits: relations are made explicit, replication and inconsistency are avoided, distant reading, that is observing the domain as a whole, is made possible. Such knowledge base allows keepers and curators to focus their efforts on collections: the available information is an useful reference for identifying pieces, for historical research and for documentation of exhibits. Sharing of information guarantees reliability of contents and reinforces cooperation among colleagues. Computing History Knowledge Base (CHKB) is a software system to maintain in a shared knowledge base all information about all collections related to a particular field of the scientific and technological heritage: computing history. CHKB permits a double economy: it is a full exploitation of available technologies and saves time and effort of keepers and curators. CHKB model can be applied to other domains of cultural heritage.

Key words:

knowledge base, technological and scientific heritage, cooperative work, computing history.

INTRODUZIONE

La cosiddetta rivoluzione digitale ha cambiato molto il modo in cui si concepisce e si tratta l'informazione. La scienza ci ha dato i concetti per modellare le relazioni tra le informazioni e la conoscenza che rappresentano. Oggi abbiamo anche gli strumenti tecnologici per trattare in modo efficiente la conoscenza – non solo i dati. Sono anche aumentati i canali d'accesso: Internet è una realtà che rende più facile la condivisione e la collaborazione fra persone.

Computing History Knowledge Base (CHKB) sfrutta questa disponibilità di tecnologie a vantaggio della catalogazione. CHKB riunisce in un'unica base di conoscenza (Knowledge Base, KB) tutte le informazioni relative a tutti i pezzi di tutte le collezioni afferenti a un settore particolare del patrimonio scientifico: la storia dell'informatica (Cignoni, 2015).

L'articolo descrive come CHKB sia un diverso approccio alla catalogazione, capace di usare meglio il tempo di conservatori e catalogatori attraverso l'uso di una KB strutturata, condivisa e autorevole. Nel

prossimo paragrafo si discute delle principali differenze con i cataloghi tradizionali. Nel paragrafo successivo sono descritte brevemente le funzionalità del prototipo realizzato. Un altro paragrafo presenta gli sviluppi futuri.

UN NUOVO APPROCCIO ALLA CATALOGAZIONE

I cataloghi tradizionali rappresentano le collezioni come semplici elenchi di pezzi, ma "una scheda, un pezzo" è un'impostazione che risale a quando l'unica tecnologia disponibile per trattare le informazioni era la carta. Come struttura dati è solo una tabella che (forse) contiene, ma non rende facilmente utilizzabili, le relazioni tra i reperti in catalogo. Anche quando informatizzati, siffatti cataloghi perpetuano la struttura "cartacea" di figura 1: meri elenchi, separati fra loro, spesso con formati diversi.

Le proposte di formati standard come le schede ICCD PST (ICCD, 2014) mantengono l'impostazione tradizionale aumentando solo il numero di campi. Ma, dato un dominio specifico, questa impostazione aumenta solo il numero di "N/A" rimanendo insufficiente a catturare la conoscenza esistente sui pezzi.

Nei cataloghi tradizionali, la scheda di un reperto replica le informazioni della "classe" di cui è istanza: della specie di un animale tassidermizzato o, nel settore del patrimonio scientifico di nostro interesse, del modello di un vecchio calcolatore. Ogni museo, nel suo catalogo, avrà una copia di tali informazioni. Ogni catalogatore replicherà lo sforzo di cercarle e registrarle, rischiando inconsistenze e imprecisioni fra i vari cataloghi. In pratica, è ignorata la possibilità di strutturare e condividere le informazioni.

CHKB distingue le informazioni su un reperto dalle informazioni sul modello, separandole rispettivamente in scheda reperto e scheda tecnica.

Esiste una scheda reperto per ogni pezzo di una collezione, ma contiene solo informazioni pertinenti alla fisicità del pezzo: lo stato di conservazione e funzio-

namento, la provenienza, la posizione in vetrina o in deposito ecc.

Le schede tecniche descrivono invece le caratteristiche di un modello comuni a tutti i pezzi, come le misure, l'anno di lancio sul mercato, la frequenza di clock del processore, la quantità di memoria e così via. Le schede tecniche possono essere collegate tra loro: un calcolatore sarà collegato al processore che utilizza, ma le informazioni sul processore sono solo nella relativa scheda tecnica (e.g. le frequenze di clock ammissibili). Particolari schede tecniche descrivono persone e aziende. Sono riferite per esplicitare le relazioni che legano i modelli alle persone che li hanno progettati e alle aziende che li hanno prodotti, le aziende ai loro fondatori. Naturalmente, una scheda tecnica è collegata a tutte le schede reperto pertinenti, esplicitando così in quali collezioni è conservato un pezzo di quel modello. In termini di gestione della conoscenza, le schede tecniche rappresentano i "pezzi" di informazione che collegano i reperti delle collezioni (fig. 2).

Le schede reperto sono responsabilità dei conservatori e dei curatori delle collezioni. Le schede tecniche sono invece uniche e condivise, i loro contenuti sono mantenuti autorevoli tramite un processo di peer review a cui può partecipare tutta la comunità di esperti.

In origine i cataloghi sono nati per gli addetti ai lavori: per conoscere la collezione di cui si è responsabili, per preparare una mostra o una sala di un museo, per fare una ricerca storica o tecnologica. Ma la conoscenza "dentro" i cataloghi oggi può essere un mezzo di restituzione al pubblico. La struttura che CHKB dà alle informazioni fa del catalogo uno strumento di narrazione: schede reperto e schede tecniche possono essere collegate in percorsi tematici. Un percorso può, banalmente, replicare un'esposizione, oppure accompagnare alla scoperta di un particolare aspetto della storia o della tecnologia dell'informatica, anche toccando pezzi conservati in musei diversi. Oppure, ancora, la rete di collegamenti può essere usata per leggere le informazioni "a distanza" (Jänicke et al., 2015), osservandole nel loro insieme e scoprendo nuove interpretazioni e narrative.

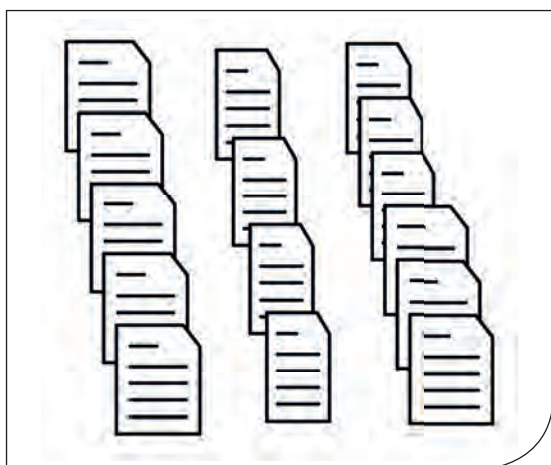


Fig. 1. Vecchi cataloghi: semplici elenchi di schede, separati e spesso con formati diversi.

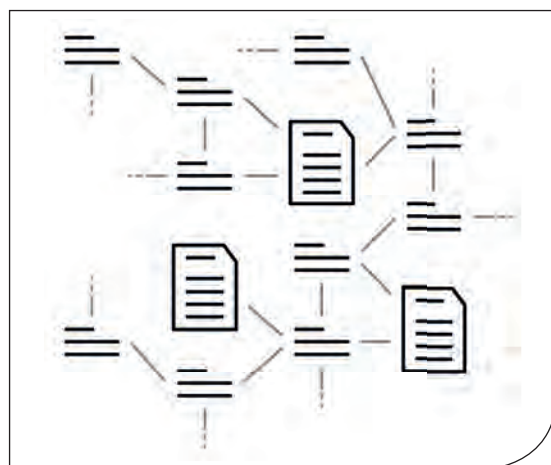


Fig. 2. Base di conoscenza: una struttura condivisa che collega tutte le informazioni di un dominio.

PROTOTIPO

Nel corso di alcune tesi di laurea triennale e magistrale all'Università di Pisa è stato realizzato un prototipo di CHKB come applicazione web. Le tecnologie usate sono: PostgreSQL per la base dati, PHP5 su server Apache per le richieste HTTP, Bootstrap, JavaScript e CSS, per la grafica. Il prototipo è disponibile su web (v. sito web 1). Il prototipo implementa i diversi tipi di scheda. Le schede reperto hanno campi specifici per tipo di reperto: hardware, software, documento, cimelio, altro. Le schede tecniche descrivono i modelli per gli stessi tipi più persone e aziende. Tutte le schede sono collegabili fra loro.

Il prototipo gestisce il processo di redazione e revisione che riguarda le schede tecniche. Anche quando un pezzo è unico (e.g. un calcolatore realizzato per ricerca) è utile mantenere la distinzione fra scheda reperto e scheda del modello: sottolinea le diverse responsabilità e competenze, la prima è redatta dai curatori/conservatori del pezzo, la seconda dalla più ampia comunità di storici. Il processo di redazione e revisione è controllato: ogni versione nella storia di una scheda è mantenuta, tracciando modifiche e autori. Anche la visibilità delle schede è controllata: i contenuti di CHKB aspirano a essere pubblici, ma per le schede reperto è possibile gestire le esigenze di riservatezza proprie di molti collezionisti privati.

Alle schede possono essere collegati contenuti digitali, sia generici (e.g. un filmato pubblicitario di un Apple) collegato alla relativa scheda tecnica sia specifici (e.g. la documentazione fotografica di un pezzo o la digitalizzazione di un manuale conservato collegati alle relative schede reperto).

Le relazioni fra le informazioni della KB sono ottenute collegando le schede. Una scheda reperto, oltre che alla scheda tecnica del modello, può essere collegata alla scheda persona del proprietario del pezzo, magari un personaggio famoso. La scheda tecnica sarà collegata alla scheda persona del progettista e alla scheda azienda della società produttrice, e così via.

Il prototipo prevede diversi tipi di utenti. L'utente base può consultare tutti i contenuti escluse le schede reperto definite riservate. Può suggerire la creazione di una scheda tecnica. Può essere promosso a ruoli più impegnativi.

Gli utenti responsabili di una collezione possono modificare le informazioni sulla collezione e attribuire ad altri utenti il ruolo di catalogatori. I catalogatori possono aggiungere nuove schede reperto alla collezione o modificare quelle esistenti.

Gli utenti revisori fanno parte del gruppo di esperti che garantisce l'autorevolezza della base di dati. Possono creare schede tecniche, eventualmente accettando proposte degli utenti. Possono attribuire il ruolo di redattori assegnando la redazione o la revisione delle schede tecniche.

Gli utenti narratori possono creare dei percorsi tematici collegando schede tecniche, schede reperto e contenuti digitali per raccontare una storia. Il prototipo riconosce l'intersezione dei percorsi, segnalando all'utente gli incroci nelle narrazioni della storia dell'informatica.

CONCLUSIONI E SVILUPPI FUTURI

Le tecnologie alla base di CHKB sono disponibili: non utilizzandole si perde un'opportunità. Condividendo la conoscenza, conservatori e curatori sono liberi di concentrarsi sulle "loro" collezioni e sulla loro attività di ricerca. Condivisione e peer review, oltre a garantire l'autorevolezza dei contenuti, favoriscono quella collaborazione che, storicamente, è un motore del progresso scientifico (Vermeulen et al., 2013). La possibilità di coinvolgere gli utenti nella redazione dei contenuti permette l'inclusione di attori importanti quali i collezionisti privati. La storia dell'informatica è un caso di studio, ma il modello di CHKB può essere applicato ad altri campi del patrimonio culturale e tecnico-scientifico. Considerando la quantità di informazioni sulla storia dell'informatica disponibili su internet, uno sviluppo futuro di CHKB sarà l'utilizzo di ricerche automatizzate che usano l'ontologia della KB e gestiscono la potenziale inaffidabilità delle informazioni recuperate (Dou et al., 2015).

RINGRAZIAMENTI

Un ringraziamento particolare va a G. Lettieri dell'Università di Pisa, relatore delle tesi che hanno realizzato il prototipo di CHKB. Come sviluppo dell'idea di CHKB è in preparazione una proposta di progetto europeo (Cignoni & Cossu, 2016) per una base di conoscenza internazionale sulla storia dell'informatica, per i molti scambi di idee un ringraziamento va a G.A. Cossu, N. Zanetti e S. Salvadori di Hyperborea srl e a F. Niccolucci dell'Università di Firenze.

BIBLIOGRAFIA

- CIGNONI G.A., 2015. *CHKB: dare struttura (visitabile) alle collezioni tecnico-scientifiche*. Presentazione alla IV Conferenza Nazionale dell'Associazione per l'Informatica Umanistica e Cultura Digitale, Torino, 17-19 dic. 2015.
- CIGNONI G.A., COSSU G.A., 2016. The Global Virtual Museum of Information Science & Technology, a Project Idea. In: Atti della Conferenza IFIP WG 9.7, New York. *IFIP Advances in Information and Communication Technologies*, 491: 101-114, Springer.
- DOU D., WANG H., LIU H., 2015. *Semantic Data Mining: A Survey of Ontology-based Approaches*. In: Atti 9th IEEE International Conference on Semantic Computing (USA). IEEE.
- ICCD, 2014. *Normativa PST - Patrimonio Scientifico e Tecnologico*, v. 3.01.
- JÄNICKE S., FRANZINI G., CHEEMA M.F., SCHEUERMANN G., 2015. *On Close and Distant Reading in Digital Humanities: A Survey and Future Challenges*. In: Borgo R., Ganovelli F., Viola I. (eds), Eurographics Conference on Visualization (EuroVis) (2015). Atti di EuroVis 2015. STAR – State of The Art Report.
- VERMEULEN N., PARKER J.N., PENDERS B., 2013. Understanding life together: A brief History of collaboration in biology. *Endeavour*, 37(3): 162-171.

Siti web (ultimo accesso 26.02.2017)

- 1) Prototipo CHKB – <http://hmr.di.unipi.it/CHKB.html>