

Building Smart City Views through Web Information Clouding



Stefano Montanelli
Silvana Castano
Lorenzo Genta

Congresso Nazionale AICA 2011
Torino, 15-17 Novembre 2011



Regione Lombardia



*Dipartimento di Informatica e Comunicazione
Università degli Studi di Milano*

- Le recenti innovazioni web hanno radicalmente mutato i meccanismi di fruizione dei contenuti pubblicate in rete
 - *Web 2.0*: disponibilità di informazioni spontanee, definite dagli utenti e rapidamente obsolescenti
 - *Semantic web*: affermazione di una logica orientata ai dati (e non alle risorse nel loro insieme)
 - *Social web*: bisogno di partecipazione, di personalizzazione degli utenti e dei relativi interessi

Le attuali tecniche di ricerca e classificazione di risorse web sono inadeguate in scenari così fortemente dinamici e pervasivi



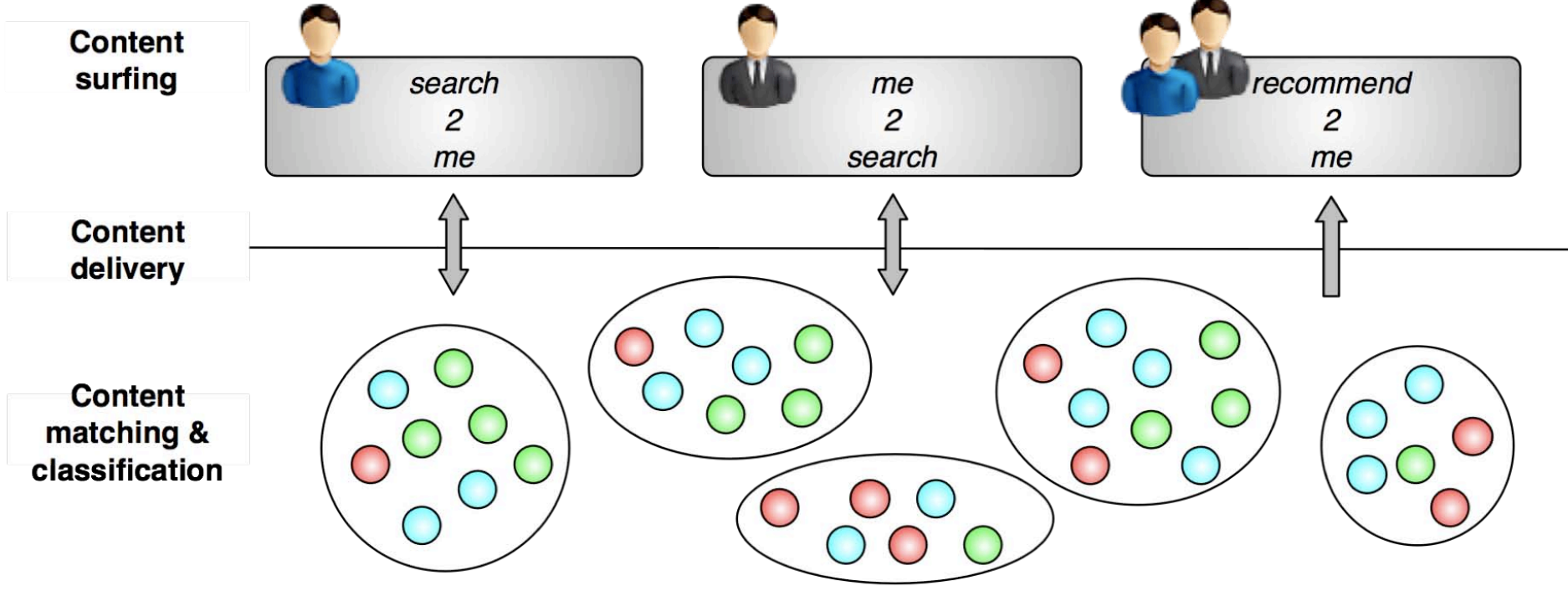


Obiettivi di Mi-Search

- Mi-Search è un progetto mirato allo sviluppo di servizi *knowledge-intensive* per la fruizione personalizzata e immediata sul territorio di contenuti web correlati
 - Sostenere la diffusione mirata di eventi e iniziative
 - Favorire l'incontro fra domanda e offerta
 - Incentivare la partecipazione attiva degli utenti
- Lo scenario considerato è il territorio milanese e il relativo patrimonio artistico e culturale



L'approccio Mi-Search



Source webs

bookmarking/
annotation
systems

delicious flickr

Tagged resources

microblogging/
social networks

twitter facebook news feed

Microdata resources

RDF(S)/OWL
repositories

DBpedia Freebase

Semantic web resources



Modello *search-2-me*



Ho in programma di visitare una mostra su De André allestita alla Rotonda della Besana. Inserisco un appuntamento in agenda



Tramite *Mi-Search* sul mio dispositivo mobile ottengo informazioni correlate all'evento



Dove si trova la Rotonda della Besana e come posso raggiungerla



Informazioni sulla mostra, su De André e sulla sua produzione artistica



Modello *search-2-me*



Ho in programma di visitare una mostra su De André allestita alla Rotonda della Besana. Inserisco un appuntamento in agenda



Tramite *Mi-Search* sul mio dispositivo mobile ottengo informazioni correlate all'evento



I commenti di altri visitatori che hanno già visitato la mostra



Utenti con interessi simili (o amici) che hanno in programma di visitare la mostra



Altri eventi simili in programma in città, offerte e promozioni per i visitatori



Modello *me-2-search*



Voglio aumentare la visibilità del mio locale di sushi in Viale Monte Nero



Cerco un'ispirazione su *Mi-Search*
Il sistema restituisce una lista di risultati



Eventi in programma nelle vicinanze



Utenti appassionati di cibo orientale



Decido di proporre un menù a prezzo ridotto per i visitatori della mostra su De André



Modello *recommend-2-me*



Lavoro in Corso di Porta Vittoria e sto tornando a casa al termine della giornata



Mi-Search segnala un ristorante sushi (di cui sono appassionato) e una promozione per i visitatori della mostra su De André



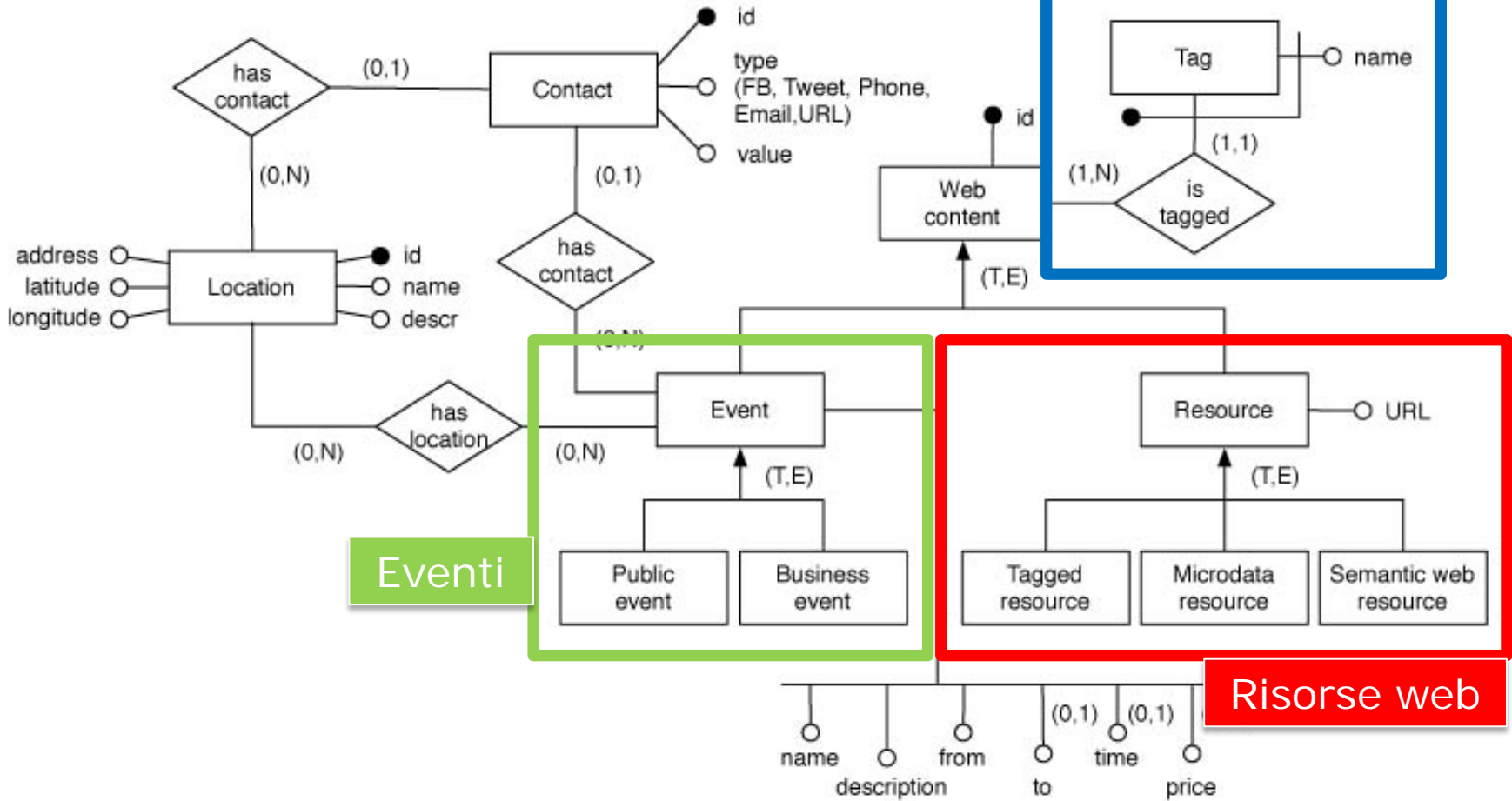
Raccolgo informazioni sulla mostra e decido di visitarla. In seguito andrò al ristorante sushi



Al termine della serata scrivo un commento: la mostra è bella ma sono necessarie almeno 2 ore per la visita, il ristorante è stata una piacevole scoperta



II repository Mi-Search-DB



- *Periodicamente* i contenuti web memorizzati in Mi-Search-DB vengono confrontati mediante *tecniche di similarità terminologica* basate sull'uso degli insiemi di tag relativi ai contenuti memorizzati

$$\text{sim}(r_1, r_2) = \frac{2 |\{t_x \sim t_y\}|}{|T(r_1)| + |T(r_2)|} \quad \begin{array}{l} t_x \in T(r_1) \\ t_y \in T(r_2) \end{array}$$

r_1, r_2 sono i due contenuti confrontati,
 T è il term equipment di un contenuto web

- Le operazioni di matching sono precedute da operazioni di tokenizzazione dei termini composti, lemmatizzazione e stemming



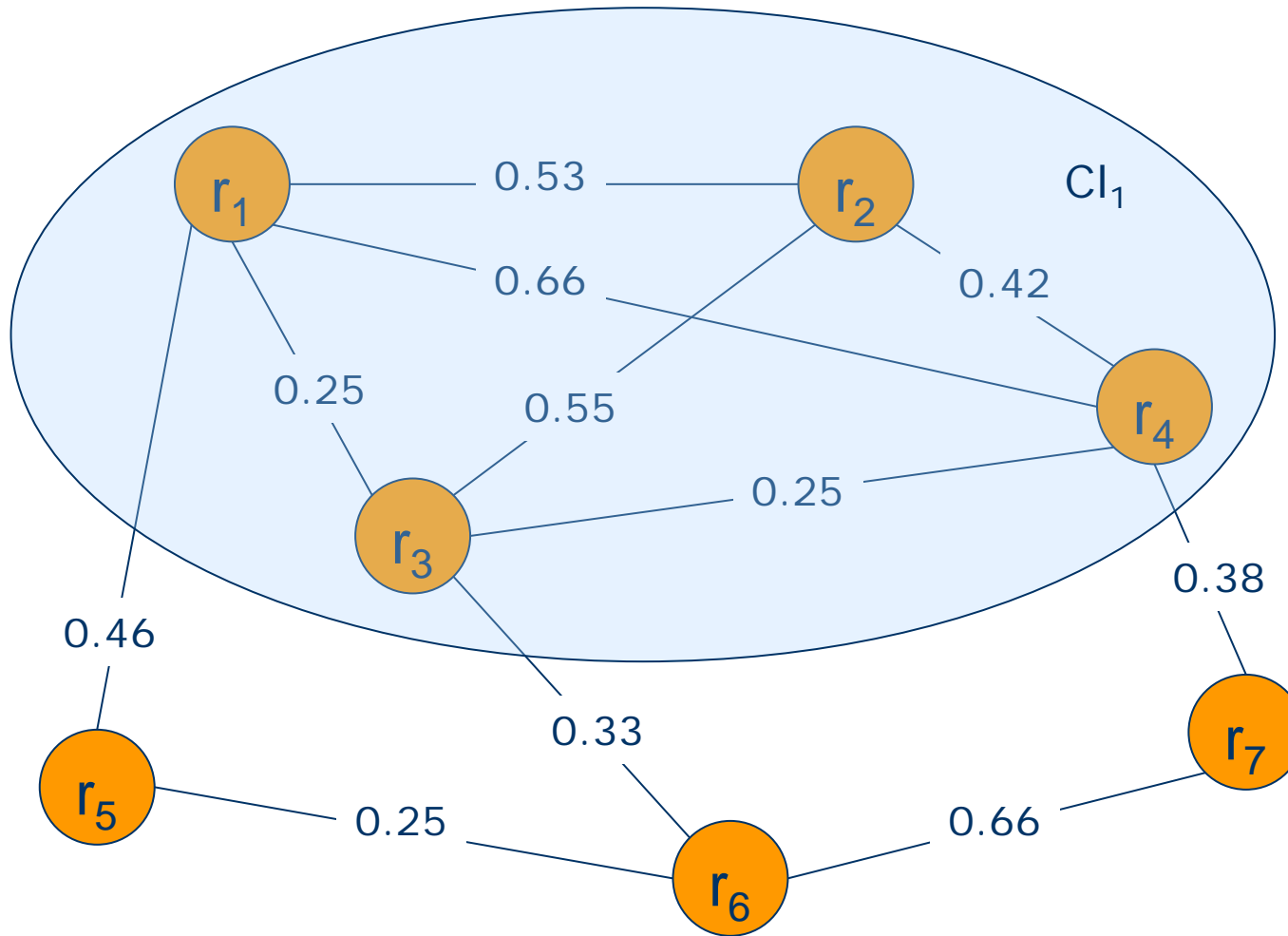


Content clustering

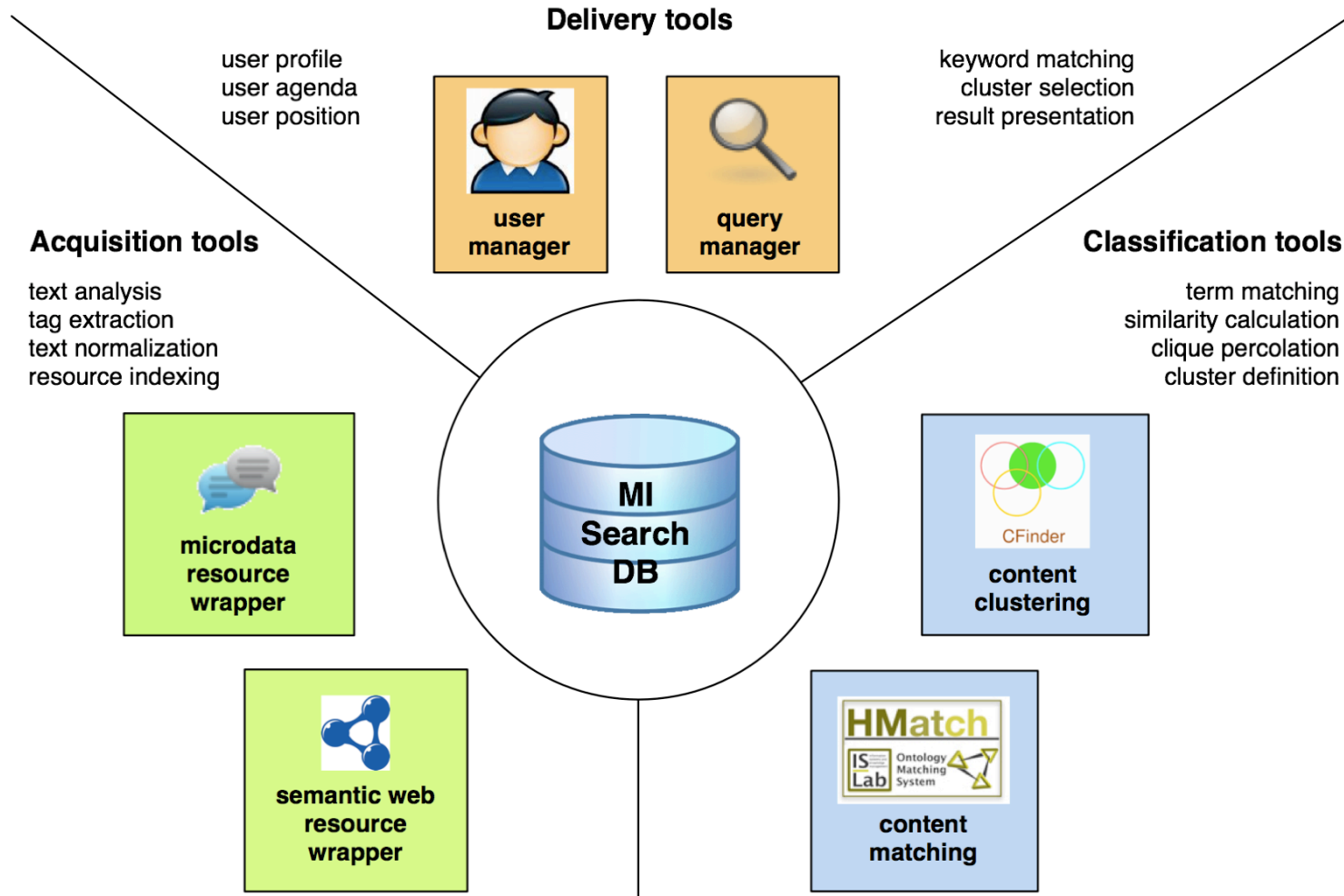
- I risultati prodotti dalle tecniche di matching generano un *grafo* in cui i nodi sono i contenuti web e gli archi stabiliscono connessioni fra i contenuti simili etichettandoli con i valori di similarità
- Le tecniche di clustering producono un insieme di *cluster di contenuti simili* identificando i gruppi di nodi che con il maggior livello di coesione (tecnica di clique percolation)



Esempio di clustering



Architettura di Mi-Search



- **Tecniche di classificazione**
 - Definizione delle tecniche di clustering (completamento)
 - Realizzazione di strumenti prototipali per *content matching* e *content clustering*
- **Caso di studio**
 - Esecuzione di test sui contenuti acquisiti
 - Analisi dei risultati
- **Architettura di Mi-Search**
 - Definizione dei requisiti relativi agli scenari d'uso (search-2-me, me-2-search, recommend-2-me)
 - Progettazione dello schema (modello definitivo)

