

Integrazione dei Dati su Web

9 CFU - A.A. 2017/18

Docente: Prof. Gennaro Costagliola

Indicazioni per lo sviluppo e documentazione del progetto

La documentazione del progetto dovrà contenere le seguenti informazioni:

1. Specifica chiara del problema e descrizione funzionale dell'applicazione finale prodotta.
2. Presentazione dell'architettura: Datawarehouse, Mediator oppure ibrida e l'architettura software del proprio sistema. Indicare chi ha lavorato su cosa. Utilizzare comunque un database/repository come cache per i dati estratti dal web inserendo un timestamp per ciascun dato memorizzato e prevedendo il suo aggiornamento in base al grado di volatilità. Il database deve sopravvivere alla chiusura delle sessioni (livello applicazione).
3. Numero e tipo delle fonti: sito web, API, DB, documenti pdf, o altro. Indicare inoltre il grado di volatilità dei dati (informazioni statiche, con cambi periodici, altamente volatili, etc.). Il progetto dovrà contenere **almeno 2** fonti a testa per ciascun partecipante al progetto (due studenti: 4 fonti, tre studenti: 6 fonti, etc.).
4. Nell'applicazione si dovranno utilizzare i dati di almeno una fonte per recuperare dati da un'altra fonte (joint).
5. Descrizione tramite datalog dello schema locale dei dati che si intende utilizzare per ciascuna fonte. ← Controllare se è tutto corretto
6. Descrizione di ciascun wrapper sviluppato: per una pagina web dare le espressioni Xpath (o CSS Path) utilizzate per l'estrazione dei dati, per una API descrivere i prototipi delle funzioni utilizzate, (in generale: <tipo fonte> → <descrizione formale per l'estrazione dati>).
7. Definizione dello schema globale tramite datalog del proprio sistema di integrazione.
8. Mostrare corrispondenze LAV e GAV in datalog fra lo schema globale e quelli locali.
9. Descrizione in datalog ed SQL (entrambi) delle query sullo schema globale che si intendono implementare. (Devono essere almeno due)
10. Elenco di tutte le tecnologie e formalismi utilizzati (ad es. Php, Java, API Google, XPath, JSON, XML o altro).
11. La documentazione va discussa tramite una presentazione Powerpoint-like.

Andrà consegnata inoltre una pagina web statica (senza scroll) contenente:

1. Dati del corso IDW (A.A., nome corso, etc.)
2. Nomi e riferimenti web (se presenti) degli autori del progetto
3. Breve descrizione del progetto (fatta in modo da invogliare il lettore ad usare il vostro sistema: messaggio pubblicitario)
4. Link alla presentazione powerpoint o ad eventuale altra documentazione.

5. Video demo del sistema di non più di due minuti (con descrizioni vocali o testuali).
6. Eventuale link al sistema, se pubblicato online (se le fonti sono open o se sono stati acquisiti i diritti di uso pubblico dai siti utilizzati come fonti).

La pagina web verrà pubblicata sul sito del Laboratorio CLUELAB e permetterà agli studenti del corso del prossimo anno di dare la loro preferenza al progetto che reputano migliore.

A handwritten signature in red ink, consisting of a stylized 'V' followed by a circle and a 'V'.