

1. Record Nr.	TD20029322
Autore	Granata, Luigi
Titolo	Advanced Techniques and Systems for Data and Process Management [Tesi di dottorato]
Altri autori (Persone)	Greco, Gianluigi Leone, Nicola
Lingua di pubblicazione	Inglese
Formato	Tesi di dottorato
Livello bibliografico	Monografia
Note	In relazione con INF/01;
Sommario	<p>Dottorato di Ricerca in Matematica ed Informatica, XXI Ciclo, a.a. 2008-2009 Il lavoro di tesi viene suddiviso in due parti che trattano rispettivamente di tecniche avanzate e sistemi per la gestione dei dati e per il mining dei processi. Sono state affrontate problematiche relative all'efficienza della risposta alle interrogazioni su una base di dati, l'integrazioni di più sorgenti di dati, e la progettazione di un sistema per il mining di processi. In particolare, i principali contributi della tesi sono: (1) Un nuovo modo di calcolare alberi di decomposizione di una interrogazione i cui query plan garantiscano un tempo di esecuzione al più di complessità polinomiale. (2) Lo studio di tecniche e metodologie innovative, basate su logica computazionale, per i sistemi per l'integrazione di sorgenti informative e lo sviluppo di un prototipo che le implementi. (3) Lo studio di tecniche ed algoritmi per il mining di processi e lo sviluppo di una suite che le implementi. (1) Tecniche di risposta alle interrogazioni su basi di dati Rispondere ad interrogazioni su una base di dati può essere un processo molto costoso da un punto di vista computazionale. Per far fronte a questa problematica, in letteratura sono stati proposti vari approcci. Alcuni di essi sono basati su moduli per l'ottimizzazione delle interrogazioni che sfruttino le informazioni quantitative e statistiche sull'istanza della</p>

base di dati, mentre altre tecniche sfruttano le proprietà strutturali degli ipergrafi delle interrogazioni. I nostri sforzi si sono rivolti in quest'ultima direzione estendendo il metodo di hypertree decomposition, considerato al momento il più potente tra quelli strutturali. Questa nuova versione, chiamata query-oriented hypertree decomposition, mira a gestire esplicitamente le variabili di output e gli operatori aggregati. Basandoci su queste nozioni, è stato implementato un ottimizzatore ibrido. Esso può essere utilizzato dai DBMS correntemente disponibili per poter calcolare i piani di esecuzione per le interrogazioni. Tale prototipo è stato integrato nel noto DBMS open source PostgreSQL. In fine questa estensione è stata validata attraverso una intensa fase sperimentale, portata avanti con PostgreSQL ed un noto DBMS commerciale, che mostra come entrambi i sistemi migliorino significativamente le loro prestazioni utilizzando le hypertree decomposition per l'ottimizzazione delle interrogazioni. 3 (2)

Tecniche per l'integrazione di sorgenti informative Per integrazione di informazioni si intende il problema di combinare i dati residenti in varie sorgenti informative, fornendo agli utenti una vista unificata di questi dati, chiamata global schema. Il nostro lavoro è stato svolto all'interno del progetto INFOMIX. Il suo scopo principale è stato quello di fornire tecniche avanzate e metodologie innovative per gli information integration systems. In breve, il progetto ha sviluppato una teoria, comprendente un modello esauriente ed algoritmi per l'integrazione delle informazioni ed l'implementazione di un prototipo di un sistema knowledge based tramite l'utilizzo della logica computazionale che integri i risultati della ricerca sull'acquisizione e la trasformazione dei dati. Un'attenzione speciale è stata dedicata alla definizione di un meccanismo per l'interazione dichiarativa da parte dell'utente e alle tecniche per la gestione di dati semistrutturati e sorgenti di dati incomplete o inconsistenti. (3)

Tecniche per il mining di processi Nel contesto della enterprise automation, il process mining è recentemente emerso come uno strumento utilissimo per l'analisi e la progettazione di processi di business complessi. Lo scenario tipico per il process mining è dato da un insieme di tracce che registrano, tramite un sistema transazionale, le attività svolte durante più esecuzioni di un processo e dall'obiettivo ricavare in maniera (semi)automatica un modello che possa spiegare tutti gli episodi registrati nelle tracce. Noi abbiamo sviluppato una Suite per le applicazioni del process mining con un'architettura aperta ed estendibile che introduce tre elementi innovativi per soddisfare i desiderata di flessibilità e scalabilità che sorgono negli scenari industriali attuali. • Il concetto di "flusso di mining", i.e., essa permette di specificare delle catene di mining complesse basate sulla connessione di task elementari. • La costruzione di applicazioni interattive basate sulla possibilità di personalizzare tipi di dati, algoritmi e l'interfaccia grafica utilizzata per l'analisi. • Scalabilità su grandi moli di