

## **Refactoring del data warehouse *GFINDER* e delle procedure per l'aggiornamento delle informazioni genomiche contenute**

*GFINDER* Genome Function Information Discoverer (<http://www.bioinformatics.polimi.it/GFINDER/>) è un sistema web multi database per l'efficace utilizzo dei dati genomici disponibili in molte banche dati eterogenee accessibili via Internet (principali pubblicazioni relative disponibili all'indirizzo <http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?CMD=search&DB=pubmed&term=GFINDER>).

*GFINDER* permette di realizzare analisi esplorative e statistiche di annotazioni funzionali e fenotipiche di gruppi di geni. Il sistema è composto da vari moduli, ognuno dedicato ad uno specifico tipo di annotazioni genetiche (processi biologici, funzioni molecolari, componenti cellulari, patologie e fenotipi patologici genetici, pathway biomolecolari, domini e famiglie proteiche) che sono raccolte in alcuni database relazionali locali. Tali annotazioni, anche espresse mediante vocabolari controllati ed ontologie, sono liberamente disponibili presso numerose banche dati remote, da dove vengono periodicamente scaricate semi-automaticamente per mantenere aggiornati i database locali del sistema. Partendo da quanto precedentemente sviluppato (database locali e procedure Java e SQL per la loro creazione e aggiornamento), la Tesi proposta dovrà realizzare il refactoring degli schemi dei database e delle procedure Java. Nel caso dei database il refactoring dovrà essere volto a produrre schemi che permettano la realizzazione efficiente delle tipiche operazioni che si realizzano sui dati di un data warehouse (slice-and-dice, roll-up, e drill-in), e quindi dell'interrogazioni dei dati contenuti. Il refactoring delle procedure Java, basato su criteri tipici dell'ingegneria del software, dovrà invece produrre un software modulare che permetta un efficiente aggiornamento dei dati nei database ristrutturati e facilmente espandibile con nuovi moduli dedicati all'integrazione ulteriori informazioni genomiche specifiche disponibili in rete.

Per informazioni:

Alessandro Campi [campi@elet.polimi.it](mailto:campi@elet.polimi.it)

Marco Masseroli [masseroli@elet.polimi.it](mailto:masseroli@elet.polimi.it)