

## Stemma codicum: analisi e generazione semi-automatica

Gioele Barabucci, [barabucc@cs.unibo.it](mailto:barabucc@cs.unibo.it)

Angelo Di Iorio, [diiorio@cs.unibo.it](mailto:diiorio@cs.unibo.it)

Fabio Vitali, [fabio@cs.unibo.it](mailto:fabio@cs.unibo.it)

Università di Bologna

La costruzione di uno stemma codicum è un processo complesso e laborioso, che richiede sia competenze e abilità intellettuali che un lungo e minuzioso lavoro manuale.

L'uso di strumenti informatici per la generazione automatica o semi-automatica degli stemma può essere un virtuoso esempio di collaborazione inter-disciplinare. Il nostro gruppo propone un approccio innovativo per lo studio degli stemma su documenti codificati in formato elettronico, ad es. TEI. Le idee alla base di questo approccio sono già state applicate con profitto al calcolo delle differenze e degli emendamenti tra documenti legislativi ma è altrettanto interessante studiarne applicabilità, potenzialità e limiti nel campo degli studi umanistici.

L'idea di base è usare algoritmi di calcolo automatico delle differenze tra documenti XML per individuare gli errori presenti tra i vari testimoni e inferire da questi uno o più possibili stemma.

L'aspetto cruciale e più innovativo è la netta separazione tra lo studio e la classificazione degli errori – e la conseguente descrizione in termini dichiarativi – e la ricerca 'meccanica' delle differenze.

In questo nuovo scenario, infatti, lo studioso di filologia si occupa della descrizione dei cambiamenti rintracciabili tra lezioni e produce una tassonomia di cambiamenti, di regole di inferenza e di importanza dei cambiamenti. Queste informazioni sono caricate in un ambiente di lavoro che lo studioso usa per generare lo stemma, vagliare diverse ipotesi di lavoro, selezionare gli errori più importanti e filtrare le informazioni non necessarie.

Un modello preliminare per descrivere gli errori ed esprimere tutte le informazioni necessarie alla generazione dello stemma codicum è quindi composto di tre parti:

- Classificazione degli errori, cioè individuazione dei possibili cambiamenti (ad es. aggiunta, permutazione) e del pattern associato ad ogni cambiamento (ad es. una permutazione è individuata dal pattern  $aXYb \Rightarrow aYXb$ ).
- Classificazione degli effetti di questi errori sulla stemma, cioè distinzione tra errori congiuntivi e disgiuntivi, e identificazione di ulteriori caratterizzazioni degli errori e del loro significato
- Regole di filtro e importanza dei cambiamenti, cioè attribuzione di diversi pesi ai cambiamenti in fase di generazione/analisi dello stemma (ad es. le modifiche sulla punteggiatura sono meno rilevanti rispetto alle modifiche sostanziali sulle parole).

Il nostro gruppo sta lavorando a un algoritmo per il calcolo delle differenze particolarmente adatto a questo contesto. L'algoritmo JNDiff, di cui esiste un'implementazione open-source, riconosce un vasto numero di operazioni e produce ottimi risultati sui documenti letterari. Caratteristica principale dell'algoritmo è la possibilità di aggregare cambiamenti atomici (inserimenti, cancellazioni) in cambiamenti complessi e più significativi per l'utente (es. spostamento di paragrafo). Un possibile sviluppo consiste nel personalizzare JNDiff per riconoscere nuovi tipi di cambiamenti descritti attraverso le tassonomie illustrate sopra, fornendo così un preciso strumento di analisi degli errori.

In conclusione, il nostro approccio demanda a una applicazione il calcolo delle differenze e l'analisi sintattica dei testi, permettendo allo studioso di focalizzarsi sulla parte filologica, dandogli la possibilità di definire le proprie tecniche in aggiunta a quelle esistenti e di sperimentare, senza sforzi manuali per la riscrittura dello stemma, varie ipotesi di lavoro durante lo sviluppo di una edizione critica.