

GlozzQL : un langage de requêtes incrémental pour les textes annotés

Yann Mathet¹, Antoine Widlöcher¹

(1) GREYC, UMR CNRS 6072, Université de Caen, 14032 Caen Cedex
{prenom.nom}@unicaen.fr

GlozzQL est un langage de requêtes couplé à un moteur d'interrogation dédié et intégré à la plateforme d'annotation Glozz. Ses deux points forts sont : (1) son immersion au coeur du processus d'annotation, permettant de requêter en même temps que l'on annote, et vice-versa (la figure 2 montre comment on peut naviguer entre les résultats de requêtes et les annotations du texte en deux clics successifs) ; (2) son caractère incrémental, au sens où des requêtes peuvent être construites sur la base d'autres requêtes (on cherche les annotations répondant à telle contrainte parmi celles répondant déjà à telle autre contrainte).

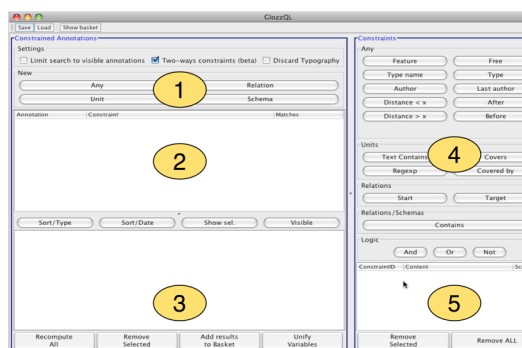


figure 1

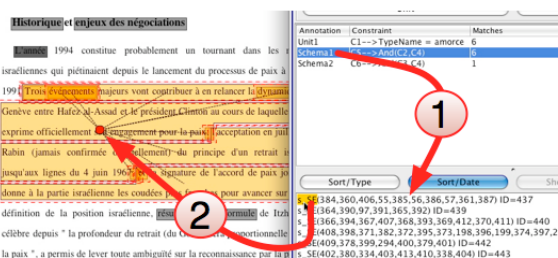


figure 2

Ce langage repose sur deux concepts inter-dépendants, les **Constraints**, définissant chacune une condition qu'une annotation doit vérifier pour être sélectionnée, et les **ConstrainedAnnotations**, ensemble d'annotations satisfaisant une contrainte donnée. C'est parce qu'une Constraint peut elle-même s'appuyer sur une ConstrainedAnnotation que des constructions incrémentales peuvent être peu à peu constituées, et produire ainsi un pouvoir expressif fort. S'adressant à un public large, ce langage est intégré sous forme **entièrement graphique** au sein de Glozz, permettant de constituer des requêtes aussi complexes que souhaité à la souris, comme illustré en figure 1 : les fenêtre 1 et 4 permettent respectivement de créer des ConstrainedAnnotations et des Constraints, lesquelles apparaissent respectivement en 2 et 5, tandis que l'ensemble des résultats relatifs à une ConstrainedAnnotation de 2 apparaît au-dessous en fenêtre 3.

La démonstration abordera GlozzQL de façon pratique et s'appuiera sur 3 exemples concrets de constitution de requêtes. Le premier, la simple détection d'annotations disposant d'une valeur de trait particulière, à savoir genre=féminin. Le second mettra en oeuvre la constitution progressive de contraintes complexes, comme la recherche d'une annotation constituée d'une relation allant d'une annotation de genre=féminin à une annotation de genre=masculin. Le troisième permettra d'entrevoir des concepts avancés de Glozz tels que l'**unification**, permettant de considérer les ConstrainedAnnotations comme des variables d'un système d'équation, et ainsi de pouvoir détecter des configuration complexes telles que des relations en triangle, ou des relations orientées dans le sens inverse du sens de lecture. Enfin, la notion de **panier** permettant de collecter sélectivement des résultats de requêtes afin soit de les enregistrer, soit au contraire de les retirer du texte annoté, sera illustrée à partir des résultats précédents.

Références

MATHET Y., WIDLÖCHER A. (2011), Stratégie d'exploration de corpus multi-annotés avec GlozzQL. *TALN 2011*, à paraître (soumission 186).