

## TER L3

### Découverte de clés dans des bases de connaissances

encadrant : Michel Leclère

contact : leclere@lirmm.fr / 04 67 41 86 16

Dans le cadre de la fusion de bases de connaissances, on se pose le problème de la découverte de clés dans une base. La découverte de clés consiste, étant donné une instance de relation (c'est à dire une table d'une BD), à identifier des ensembles minimaux d'attributs qui pourraient être clés (au sens BD) de cette relation (autrement dit si ces attributs étaient déclarés comme contrainte de clé, cette instance la respecterait).

Considérons par exemple la table :

<b>Numéro</b>	<b>Auteur</b>	<b>Titre</b>	<b>Date</b>	<b>Type</b>
123	V. Hugo	Ruy Blas	1838	Théâtre
128	V. Hugo	Les misérables	1862	Roman
143	V. Hugo	L'Âne	1880	Poésie
267	E. Zola	Nana	1880	Roman
270	E. Zola	Au bonheur des dames	1883	Roman
271	E. Zola	La joie de vivre	1883	Roman
938	C. Lelouch	Les misérables	1995	Film

Dans cette table, on peut identifier les clés minimales potentielles :

- {Numéro}
- {Auteur, Titre}
- {Titre, Date}
- {Titre, Type}

De telles clés peuvent alors être utilisées comme des règles d'identification de tuples issus de bases différentes. Sur l'exemple précédent, on pourrait alors considérer que : **si deux documents ont même titre et même auteur alors il s'agit du même document.**

Différents algorithmes ont été proposés pour la découverte de telles clés. L'objectif de ce projet est d'étudier et d'implanter l'un de ces algorithmes et de l'expérimenter sur des données réelles.

### Références

- Y. Huhtala, J. Karkkainen, P. Porkka, and H. Toivonen. Tane: An efficient algorithm for discovering functional and approximate dependencies. *Comput. J.*, 42(2):100–111, 1999.
- Y. Sismanis, P. Brown, P. Haas, and B. Reinwald. Gordian: efficient and scalable discovery of composite keys. In *Proceedings of the 32nd international conference on Very large data bases*, pages 691–702. VLDB Endowment, 2006.
- Nathalie Pernelle, Fatiha Saïs, Danai Symeonidou: An automatic key discovery approach for data linking. *J. Web Sem.* 23: 16-30 (2013)
- Manuel Atencia, Jérôme David, Jérôme Euzenat: Data interlinking through robust linkkey extraction. *ECAI 2014*: 15-20