

Accumulation of Lexical Sets: Acquisition of Dictionary Resources and Production of New Lexical Sets

DOAN-NGUYEN Hai
GETA - CLIPS - IMAG
BP 53, 38041 Grenoble, France
Fax: (33) 4 76 51 44 05 - Tel: (33) 4 76 63 59 76 - E-mail: Hai.Doan-Nguyen@imag.fr

Following is the abstract written in Vietnamese:

Tóm tắt

Trong bài này, chúng tôi trình bày một công trình về tích lũy các tập hợp từ vựng, bao gồm sự thu hồi các tài nguyên từ điển và sự sản sinh các tập hợp từ vựng mới. Phương pháp thu hồi sử dụng một bộ dịch theo cú pháp phi ngữ cảnh và các kỹ thuật sửa đổi văn bản, đã chứng tỏ dễ dùng, linh hoạt và hiệu quả. Trên cơ sở phân tích các loại hình sản sinh các tập hợp từ vựng mới, chúng tôi đề ra các thao tác cơ bản và thiết lập một cơ chế hình thức giúp mô tả và thực hiện việc sản sinh. Với các phương pháp và công cụ trên, khoảng 1,7 triệu đơn vị từ vựng đã được thu hồi và sản sinh từ các nguồn từ điển thuộc nhiều loại và độ phức tạp khác nhau. Cũng trong bài này, chúng tôi đề nghị một phương cách tổ chức động và có tính tổ hợp cho các hệ thống từ vựng, dựa trên khái niệm tích lũy ảo và các cấp trừu tượng hoá của tập hợp từ vựng.

Đoàn Nguyễn Hải

Following is the abstract written in French:

Résumé

Cet article présente notre travail sur l'accumulation d'ensembles lexicaux qui comprend l'acquisition de ressources dictionnairiques et la production de nouveaux ensembles lexicaux. La méthode pour l'acquisition, utilisant un traducteur hors-contexte gouverné par la syntaxe et des techniques de modification de texte, se montre flexible, efficace et facile à utiliser. Des opérations de base sont proposées qui forment un formalisme de spécification et d'implémentation de processus de production. Environ 1,7 million unités lexicales ont été récupérées et produites à partir de dictionnaires de types et de complexités différents. Nous proposons aussi une organisation combinatoire et dynamique pour les systèmes lexicaux, basée sur la notion d'accumulation virtuelle et sur des niveaux d'abstraction d'ensembles lexicaux.