

Le DiCoInfo Visuel

Documentation logicielle

Benoit Robichaud



Cette page web est dédiée à la présentation visuelle
de certains [liens lexicaux](#) du [DiCoInfo](#)

Relations	Langue	Précision	Graphique
<input checked="" type="checkbox"/> taxonomique <input checked="" type="checkbox"/> synonymes <input checked="" type="checkbox"/> contraires <input checked="" type="checkbox"/> dérivés <input checked="" type="checkbox"/> voisins <input type="checkbox"/> actants	<input checked="" type="radio"/> français <input type="radio"/> anglais <input type="radio"/> espagnol	<input checked="" type="radio"/> contenant <input type="radio"/> débutant <input type="radio"/> exacte	<input checked="" type="radio"/> réification <input type="radio"/> fermeture <input type="radio"/> incomplets <input type="radio"/> orphelins
<input type="button" value="Tous"/> <input type="button" value="Aucun"/> <input type="button" value="Ouvrir"/> <input type="button" value="Fermer"/>	Expression cherchée <input type="text" value="disque"/> <input type="button" value="🔍"/>		

Observatoire de Linguistique Sens-Texte (OLST)

Mars 2013

Table des matières

Introduction.....	3
Guide d'utilisation	5
Fonctionnement en bref.....	6
Paramétrage des recherches	6
Présentations des résultats.....	9
Pour en savoir plus.....	11

Introduction

Le *DiCoInfo Visuel* est une interface de « visualisation graphique » au dictionnaire terminologique *DiCoInfo*, développé sous la direction de Marie-Claude L'Homme depuis plusieurs années à l'Observatoire de linguistique Sens-Texte (OLST), Université de Montréal. Son objectif principal est de présenter aux utilisateurs une vue d'ensemble des « relations lexicales » décrites dans une ou plusieurs entrées du dictionnaire sous la forme d'un réseau. Il donne ainsi un point de vue unique sur les relations entre les termes qui y sont décrits, et il constitue de plus un outil de validation aux personnes qui rédigent et révisent ces mêmes entrées.

Tout comme l'interface de recherche principale du *DiCoInfo*, celle du *DiCoInfo Visuel* permet de faire des recherches dans les entrées de la base de données XML du *DiCoInfo* pour trouver des informations relatives à un ou plusieurs termes, mais, contrairement à la première qui affiche un grand nombre de relations lexicales incluant les liens paradigmatiques et syntagmatiques, la présente interface se concentre sur un sous-ensemble des relations lexicales qui y sont décrites (la plupart des relations paradigmatiques), et les présente spécifiquement sous la forme de *graphes* (ou *réseaux*) comme celui de la figure 1 plus bas.

Le présent document présente et décrit le *DiCoInfo Visuel*. Il se divise en trois parties qui abordent respectivement les composantes et le fonctionnement général du logiciel, les différentes options du menu qui permettent de paramétrer les recherches, et enfin la présentation des résultats aux utilisateurs ainsi que les possibilités de *navigation* données par les hyperliens insérés dans les résultats. Comme deux versions du *DiCoInfo Visuel* ont été programmées entre 2011 et 2012, l'une en PHP et l'autre en Prolog, à quelques endroits nous décrivons les différences et expliquons l'impact de quelques modifications. Le *DiCoInfo Visuel* est accessible avec un navigateur Internet standard et ne requiert aucune installation de logiciels ou bibliothèques de fonctions supplémentaires de la part des utilisateurs.

Guide d'utilisation

Le *DiCoInfo Visuel* est une interface d'interrogation des relations lexicales décrites dans les entrées terminologiques du *DiCoInfo*. Il se présente comme un formulaire de recherche sur une page Internet accessible à l'adresse donnée ci-dessous en (1) pour la première version en PHP, et à l'adresse donnée en (2) pour la version Prolog en développement. Lors du premier accès, une page d'accueil s'affiche et présente les différentes options de recherche (avec certaines valeurs sélectionnées par défaut) montrée ci-dessous à la figure 2. Comme pour les recherches depuis l'interface d'interrogation principale du *DiCoInfo*, les recherches sont aussi faites directement dans la base de données XML, puis les résultats sont colligés et présentés sur une nouvelle page qui reprend le même formulaire de recherche en y incluant les résultats.

- 1) <http://olst.ling.umontreal.ca/dicoinfo/visuel.php>
- 2) <http://olst.ling.umontreal.ca:8000/dicoinfo/visuel/search.cgi>

**Cette page web est dédiée à la présentation visuelle
de certains liens lexicaux du DiCoInfo**

Relations recherchées <input checked="" type="checkbox"/> taxonomique <input checked="" type="checkbox"/> synonymes <input checked="" type="checkbox"/> contraires <input checked="" type="checkbox"/> dérivés <input checked="" type="checkbox"/> voisins <input type="checkbox"/> actants <input type="button" value="Tous"/> <input type="button" value="Aucun"/> <input type="button" value="Ouvrir"/> <input type="button" value="Fermer"/>	Précision <input checked="" type="radio"/> contenant <input type="radio"/> débutant <input type="radio"/> exacte	Langue <input checked="" type="radio"/> français <input type="radio"/> anglais <input type="radio"/> espagnol	Graphique <input checked="" type="radio"/> fusionné <input type="radio"/> glissière <input type="radio"/> onglets
---	--	---	---

Expression cherchée

Le *DiCoInfo Visuel* utilise les gadgets Javascript de la librairie [BlueShoes](#) et le logiciel de dessin [GraphViz](#) de AT&T.
Les utilisateurs d'Internet Explorer doivent télécharger le module [SVG Viewer](#) d'Adobe.

Conception et réalisation : [Benoit Robichaud](#)

© 2011 [OLST](#)

Figure 2 : Page d'accueil du *DiCoInfo Visuel*

Outre le formulaire de recherche et les résultats que nous abordons dans les sections suivantes, cette page affiche quelques informations sur les deux paquetages [BlueShoes](#) et [GraphViz](#) utilisés respectivement pour dessiner le menu dynamique des options de recherches et la création des graphes constituant la partie centrale des résultats présentés aux utilisateurs. Ces deux programmes sont disponibles gratuitement aux adresses suivantes :

- 3) <http://www.blueshoes.org/>
- 4) <http://www.graphviz.org/>

Fonctionnement en bref

Le *DiCoInfo Visuel* est constitué de différents programmes qui se répartissent le travail. Un premier programme « maître », rédigé dans la première version en PHP et dans la seconde en Prolog pour ses capacités mieux adaptées à manipuler les relations, coordonne l'ensemble du travail. Il génère la page *html* avec laquelle interagissent les utilisateurs. Il gère donc les options de recherche ainsi que les données reçues de la base de données. Enfin, il appelle les autres programmes subordonnés. Son fonctionnement est relativement simple et peut être décrit comme un « cycle » qui consiste : à faire afficher la page d'accueil formulaire de recherche ; à faire les requêtes à la base de données XML du *DiCoInfo* ; puis après avoir colligé et traité les résultats reçus de cette dernière, à appeler le programme *GraphViz* qui dessine le graphe des relations ; et finalement, à afficher la page *html* des résultats contenant en plus cette fois le graphe et un tableau récapitulatif où les termes trouvés sont également classés en fonction de ces mêmes relations. À partir de cette nouvelle page, les utilisateurs peuvent optionnellement modifier les options de recherche, cliquer sur l'un des nœuds pour faire une nouvelle recherche avec ce terme, réécrire une toute nouvelle recherche, ou encore faire une recherche dans la page d'interrogation principale du *DiCoInfo* pour l'un des termes trouvés.

Paramétrage des recherches

Comme le montre la figure (2) plus haut, le menu d'options de recherche consiste en quatre groupes de paramètres *Relations cherchées*, *Précision*, *Langue* et *Graphique*, ainsi qu'un champ *texte* permettant la saisie de la chaîne de caractères qui contiendra à l'expression cherchée. Le *DiCoInfo* est un dictionnaire en construction. Ainsi le *Visuel* ne retient dans ses graphes que les articles dont la rédaction est terminée.

Relations cherchées

- taxonomique
 - hyperonymes
 - hyponymes
- synonymes
 - synonymes exacts
 - quasi-synonymes
- contraires
 - antonymes
 - quasi-antonymes
 - contrastifs
 - réversifs
- dérivés
 - adjectifs
 - noms
 - verbes
- voisins
- actants

Tous Aucun

Ouvrir Fermer

Figure 3 : Options de recherche des "Relations cherchées" du *DiCoInfo Visuel*

Le premier paramètre *Relations cherchées* permet de restreindre la recherche à différents sous-ensembles des relations lexicales (pour la majorité, *paradigmatiques*) décrites dans les entrées du *DiCoInfo*. Les recherches elles-mêmes sont, d'une part, toujours faites dans la vedette des entrées terminologiques (ainsi que dans les et les synonymes puisque certains termes n'ont pas d'entrée et n'apparaissent que comme valeur de ces champs), puis en fonction des relations cherchées, comme valeur des balises « lien » décrivant les liens lexicaux des termes. Les différentes relations sont regroupées en *familles* qui correspondent aux familles des liens lexicaux décrites dans les entrées terminologiques elles-mêmes. Elles permettent de faire la recherche de *classes naturelles* de relations.

Contrairement aux autres relations, la relation *actants* est particulière puisqu'elle permet de chercher des relations de types *syntagmatiques* (ou *combinatoires*) et trouve les termes typiques associés aux « rôles » de la *structure actancielle* des entrées terminologiques qui possède une telle description.

Mentionnons que, du point de vue de la collecte des résultats, ces derniers sont considérés *neutres* quant à l'endroit où la chaîne cherchée sera trouvée : lorsque la chaîne est trouvée

dans l'identificateur d'un terme, on retient ce terme et tous les liens lexicaux pour les relations recherchées ; lorsque la chaîne est trouvée dans un lien lexical, on retient ce lien et l'identificateur du terme dans lequel ce lien lexical est décrit. Le tableau qui suit énumère les contraintes effectives exprimées dans les requêtes faites à la base de données XML :

Relation cherchée	Contraintes sur les termes et les liens lexicaux cherchés
<i>Hyperonymes</i>	La chaîne « disque » est trouvée dans les vedettes (par exemple <i>disquette 1</i>), ou dans des liens d'hyperonymie (par exemple, <i>disque 1</i> de l'entrée <i>CD 1</i>).
<i>Hyponymes</i>	La chaîne « disque » est trouvée dans les liens d'hyperonymie (par exemple, <i>disque 1</i> dans l'entrée <i>CD 1</i>) ou dans les vedettes (par exemple <i>disquette 1</i>).
<i>synonymes exacts</i>	La chaîne « disque » est trouvée dans les vedettes (par exemple <i>disque rigide 1</i>), ou dans les synonymes (par exemple, <i>partition disque</i> de l'entrée <i>partition 1</i>)
<i>quasi-synonymes</i>	La chaîne « disque » est trouvée dans les vedettes (par exemple <i>disque dur 1</i>), ou dans les quasi-synonymes (par exemple, <i>espace disque</i> de l'entrée <i>espace 1</i>)
<i>Antonymes</i>	La chaîne « abon » est trouvée dans les vedettes (par exemple <i>abonnement 1</i>), ou dans les antonymes (par exemple, <i>désabonnement</i> de l'entrée <i>abonnement 1</i>)
<i>quasi-antonymes</i>	La chaîne « enreg » est trouvée dans les vedettes (par exemple <i>enregistrer 1</i>), ou dans les quasi-antonymes (par exemple, <i>enregistrer</i> de l'entrée <i>effacer 1</i>)
<i>contrastifs</i>	La chaîne « disque » est trouvée dans des vedettes (par exemple <i>disque dur 1</i>), ou dans les quasi-synonymes (par exemple, <i>disque dur</i> de l'entrée <i>mémoire vive 1</i>)
<i>réversifs</i>	La chaîne « compil » est trouvée dans les vedettes (par exemple <i>compiler 1</i>), ou dans les réversifs (par exemple, <i>décompiler</i> de l'entrée <i>compiler 1</i>)
<i>dérivés adjectivaux</i>	La chaîne « exéc » est trouvée dans les vedettes (par exemple <i>exécutable 3</i>), ou dans les dérivés adjectivaux (par exemple, <i>exécutable 3</i> de l'entrée <i>exécuter 1a</i>)
<i>dérivés nominaux</i>	La chaîne « exéc » est trouvée dans les vedettes (par exemple <i>exécution 1a</i>), ou dans les dérivés nominaux (par exemple, <i>exécution 1a</i> de l'entrée <i>exécuter 1a</i>)
<i>dérivés verbaux</i>	La chaîne « exéc » est trouvée dans les vedettes (par exemple <i>exécuter 1a</i>), ou dans les dérivés nominaux (par exemple, <i>exécuter 1a</i> de l'entrée <i>exécution 1a</i>)
<i>Voisins</i>	La chaîne « disque » est trouvée dans les vedettes (par exemple <i>disque dur 1</i>), ou dans les voisins (par exemple, <i>disque dur</i> de l'entrée <i>clé USB 1</i>)
<i>Actants</i>	La chaîne « compil » est trouvée dans les vedettes (par exemple <i>compiler 1</i>), ou dans les actants (par exemple, <i>programmeur</i> de l'entrée <i>compiler 1</i>)

Tableau 1 : Contraintes exprimées dans les requêtes faites pour trouver les différents types de relations à la base de données XML *DiCoInfo*

Le second paramètre *Précision* présente un choix exclusif et permet de spécifier la façon dont la chaîne de caractères à chercher peut être trouvée, soit *exactement* ou *partiellement*, lors des fouilles dans les entrées terminologiques. En choisissant alternativement comme valeur « contenant », « débutant » ou « exacte », cette option permet la construction d'une *expression régulière* qui cherchera les relations entre différents sous-ensembles possible de termes, comme le montre l'exemple de recherche pour la chaîne « disque » dans le tableau 2 ci-dessous.

Précision	Termes trouvés
<i>exacte</i>	disque
<i>débutant</i>	disque disque compact disque dur disque rigide disquette
<i>contenant</i>	disque disque compact disque dur disque rigide disquette espace disque lecteur de disque partition disque répertoire disque unité de disque

Tableau 2 : Différents sous-ensembles de termes ciblés en fonction des valeurs de l'option de recherche « Précision »

Le troisième paramètre « langue » est également un choix exclusif et permet de spécifier dans quel sous-ensemble des termes français, anglais ou espagnols du *DiCoInfo* se fera la recherche.

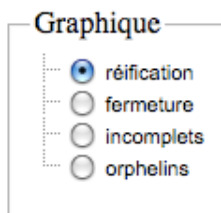


Figure 4 : Option "Graphique" du *DiCoInfo Visuel*

Le dernier paramètre *Graphique*, implémenté seulement dans la seconde version en Prolog, permet de spécifier différentes méthodes de recherche permettant soit une recherche plus exhaustive et une présentation des résultats plus précise, soit d'aider à la validation des relations entre les termes trouvés. En choisissant la valeur (par défaut) « réification », le programme passe en revue les résultats et rend *explicite* dans le graphe créé les relations cherchées qui existent entre des termes trouvés, mais qui ne contiennent ni l'un ni l'autre la chaîne recherchée. Par exemple ici entre le terme *CD₁* et les termes *clé USB₁* et *cartouche₂* :

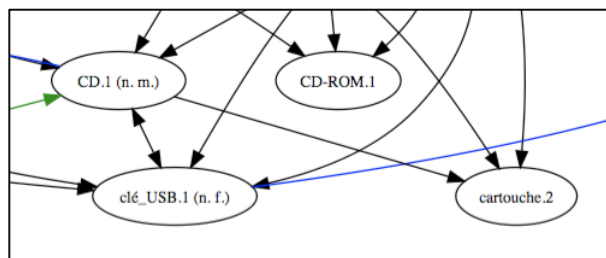


Figure 5 : Détails de la recherche de voisins contenant la chaîne « disque » dans le *DiCoInfo Visuel*

De la même façon, la valeur « fermeture » permet d'explicitier que l'utilisateur attend la *fermeture finie* du graphe pour une relation donnée. Par exemple, la recherche pour la *fermeture finie* des hyperonymes d'un terme cherchera récursivement *en remontant* toute la hiérarchie des hyperonymes même si ceux-ci ne contiennent pas la chaîne recherchée ou n'ont

pas de relation directe avec un autre terme qui la contient. Mentionnons que cette option permet en principe de mieux visualiser la place des termes ciblés, mais qu'elle peut prendre beaucoup du temps à calculer notamment lorsque plusieurs de relations sont cherchées lors d'une même requête.

Au contraire, le choix des valeurs « incomplets » et « orphelins » permet de vérifier l'intégrité de certaines relations en occultant une partie des résultats pour mieux voir certaines des relations. La valeur « incomplets » permet d'afficher les relations qui devraient être *bidirectionnelles* (donc décrites dans les deux entrées des termes en relation), mais qui ne le sont pas encore (par exemple, les relations *quasi-synonymes*, *quasi-antonymes*, *voisins*, etc.). La dernière valeur « orphelins » vise à identifier les couples de termes dont l'un des deux membres contient la chaîne recherchée, mais qui sont disjoints du reste du graphe des résultats. C'est notamment le cas lorsque la chaîne recherchée est trouvée dans un lien lexical d'un terme qui lui-même n'a aucun lien avec les autres termes contenant la chaîne recherchée (par exemple, pour la recherche de *disque*, une relation *synonyme* entre le terme *partition* et son synonyme *partition disque* est trouvée, mais aucune relation entre le terme *partition* et l'une des acceptions de *disque* n'a pas encore été décrite).

Présentations des résultats

Les résultats d'une recherche dans l'interface du *DiCoInfo Visuel* sont présentés sous deux formats. D'une part, un graphe comme celui présenté dans la figure 1 est généré par le programme de dessin *GraphViz*, puis inséré comme élément central de la page des résultats. Dans celui-ci, les termes sont représentés par les nœuds, et les relations par les arcs. D'autre part, un tableau récapitulatif classe et ordonne les termes en fonction des différentes relations pour lesquelles ils ont été retenus. Ce tableau est particulièrement utile pour vérifier rapidement si un terme a été trouvé dans l'une des relations recherchées. Nous abordons ces deux formats dans les paragraphes qui suivent.

Au moment de créer le graphe, certains détails sont spécifiés au programme *GraphViz* pour permettre aux utilisateurs une meilleure visualisation et compréhension des résultats. D'abord, les termes contenant l'expression recherchée apparaissent en gras et sont colorés en vert pour mieux les repérer parmi l'ensemble des nœuds. Ensuite, les arcs représentant les différentes relations trouvées entre les termes sont également colorés et le jeu des couleurs choisies est le même que celui utilisé dans le menu permettant sélectionner les *Relations cherchées*, ainsi que celui utilisé dans le tableau récapitulatif. Mentionnons que les différents sous-types de relations au sein des familles *synonymes* et *contraires* sont distingués par différents styles d'arcs (par exemple, les arcs des *synonymes exacts* et des *antonymes* sont pleins, alors que ceux des *quasi-synonymes* et des *quasi-antonymes* sont pointillés). De plus, les relations trouvées dans les deux descriptions des termes membres sont terminées à chaque extrémité par une *tête de flèche*. Enfin, les nœuds des termes ayant été trouvés dans l'attribut « identificateur » de balises « vocable » sont étiquetés de leur partie du discours (permettant ainsi de distinguer facilement à la fois les homographes ainsi que les sous-types de dérivés) et possèdent également un hyperlien permettant de relancer une nouvelle recherche dans le *DiCoInfo Visuel* paramétrée avec leur graphie comme nouvelle chaîne recherchée. Dans ce lot de détails, mentionnons pour terminer le cas exceptionnel des arcs reliant les actants qui contrairement aux autres types de relations sont eux étiquetés avec le nom du *rôle* qu'occupe un terme dans la structure actancielle où il a été trouvé, comme le montre la figure 6 suivante :

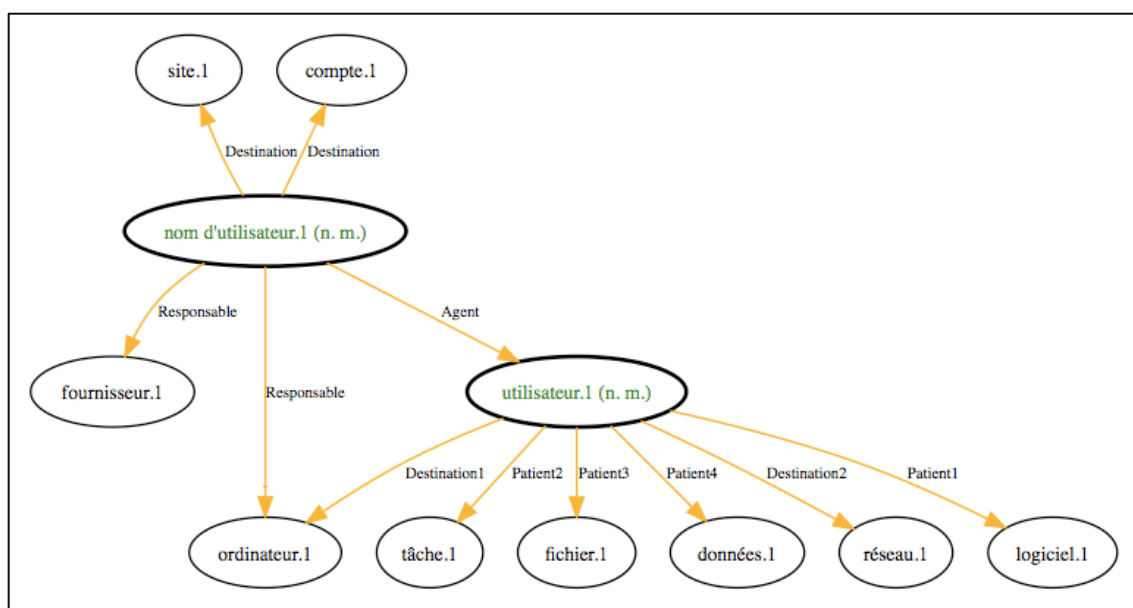


Figure 6 : Graphe des relations de type *actants* trouvées pour la chaîne « utilisateur »

Le graphe des relations est suivi au bas de la page des résultats par un tableau récapitulatif qui classe et ordonne les différents termes en fonction des relations auxquelles il participe. Ce tableau est construit par le programme principal qui n'affiche que les colonnes correspondant aux relations effectivement trouvées dans les résultats. Mentionnons que les termes impliqués dans différentes relations se retrouveront dans autant de colonnes, ce qui permet comme nous l'avons souligné plus tôt de repérer rapidement si un terme précis a été trouvé dans une relation particulière. En triant également les termes à l'intérieur de chaque colonne, il est également plus facile de repérer les fautes de frappe (orthographe et attribution d'un numéro d'acceptation erroné). Enfin, chaque terme possède un hyperlien qui permet d'un simple clic une recherche dans l'interface principale du *DiCoInfo* (Figure 7).

Liens vers les articles des lexies (6 relations)					
taxinomiques	taxinomiques	synonymes	quasi-synonymes	contrastifs	voisins
disque compact.fr disque dur.1.fr disque rigide.fr disque.1.fr disque.2.fr disquette.1.fr support de stockage.1.fr	CD.1.fr disque.2.fr	CD.1.fr disque compact.fr disque dur.1.fr disque rigide.fr	disque dur.1.fr disque.1.fr	disque dur.1.fr disque.1.fr mémoire vive.1.fr	CD-ROM.1.fr CD.1.fr cartouche.2.fr clé USB.1.fr disque dur.1.fr disque.1.fr disque.2.fr disquette.1.fr mémoire.2.fr mémoire.3.fr unité de mémoire.1.fr unité de stockage.1.fr

Figure 7 : Tableau récapitulatif des termes participants aux résultats

Pour en savoir plus

L'Homme, M.-C. (2011). *DiCoInfo, Dictionnaire fondamental de l'informatique et de l'Internet*.
<http://olst.ling.umontreal.ca/cgi-bin/dicoinfo/search.cgi>

Robichaud, B. (2012). "Logic Based Methods for Terminological Assessment", In *Language Resources and Evaluation (LREC 2012)*. Istanbul, Turquie.

Robichaud, B. (2011). "A Graph Visualization Tool for Terminology Discovery and Assessment". In Boguslavsky, I. et L. Wanner (eds.) *Actes de la 5e Conférence internationale sur la Théorie Sens-Texte (MTT'11)*, Barcelone, Espagne.