Mesures de similarité pour l'aide au consensus en anatomie pathologique

Maxime Thieu[1], Olivier Steichen[2], Eric Zapletal[1], Marie-Christine Jaulent[1] et Christel Le Bozec[1],

[1] INSERM ERM202, SPIM, UFR Broussais-Hotel-Dieu, Paris, France,
{maxime.thieu, eric.zapletal, marie-christine.jaulent, christel.lebozec}@spim.jussieu.fr
[2] Service d'Immunopathologie, Hôpital Saint-Louis, Paris, France,
ost@club-internet.fr

Résumé : Le système IDEM (Images et Diagnostics par l'Exemple en Médecine) a pour objectif d'assister les experts dans la constitution de descriptions consensuelles de cas anatomopathologiques. Ce système s'appuie sur des mesures de similarités entre les termes de pathologie tumorale mammaire organisés en réseau sémantique. Afin de pouvoir comparer plusieurs mesures de similarités et pouvoir valider l'organisation des termes, des développements ont été effectués par l'extension d'un éditeur d'ontologie. Les résultats de l'évaluation sont présentés.

Mots-clés : Ingénierie des connaissances ; réseaux sémantiques ; ontologies ; mesures de similarité sémantique ; aide au consensus

1 Introduction

La prise en charge du cancer du sein a beaucoup évolué ces dix dernières années du fait d'un diagnostic plus précoce et d'une meilleure personnalisation des traitements. Participant aux efforts de prévention, les pathologistes (médecins spécialistes qui examinent au microscope les cellules et tissus prélevés) sont amenés à reconnaître de façon plus fiable les lésions pré-cancéreuses et les lésions précoces de cancer. Malheureusement, la variabilité diagnostique entre pathologistes, largement rapportée dans la littérature, reste importante (Fleming, 1996).

Plusieurs études ont montré qu'il est possible de réduire la variabilité diagnostique entre pathologistes si ces derniers s'accordent préalablement sur un système de classification diagnostique (De Vet *et al.*, 1995). Si, de plus, les experts s'assurent lors d'une réunion de consensus autour d'un microscope multi-têtes qu'ils reconnaissent de façon reproductible des critères diagnostiques et/ou pronostiques, ils peuvent constituer des recommandations de qualité pour l'interprétation diagnostique. Dans ce cas, l'impact de la démarche de consensus sur la réduction de la variabilité diagnostique est plus important.

L'organisation de séances de consensus autour de microscopes multi-têtes est compliquée par la délocalisation et la faible disponibilité des experts pathologistes.

Le travail mené depuis quelques années autour du projet IDEM vise à développer une plateforme de téléconsensus permettant aux experts pathologistes d'accéder, de décrire et d'annoter des images histologiques de cas et de participer à la définition d'un consensus sur la base de ces annotations (Le Bozec *et al.*, 2000). La plateforme s'appuie actuellement sur les technologies liées à Internet pour l'organisation et la résolution du consensus (Zapletal *et al.*, 2003). Une tâche essentielle de l'application de téléconsensus est de simplifier la comparaison des descriptions d'images et la réalisation du consensus, en s'appuyant sur la détection automatique des éléments concordants ou discordants dans les descriptions de différents pathologistes pour fournir des « amorces » de consensus qui seront validées ou invalidées par les experts. Les étapes correspondant à la réalisation de cette tâche sont :

1. Modéliser le travail de consensus et représenter la description des cas anatomopathologiques dans un formalisme permettant leur comparaison.
2. Mettre en pratique des mesures de similarité sémantique, pour comparer les descriptions de plusieurs experts et proposer des amorces de consensus.
3. Développer un environnement logiciel permettant d'évaluer les résultats.

Le premier point a fait l'objet d'un travail spécifique publié récemment, rapidement exposé dans la partie 2, et qui a abouti à la construction d'un réseau sémantique adapté à la description et à la comparaison de descriptions d'images en pathologie tumorale mammaire (Steichen *et al.*, 2003).

Le travail présenté dans cet article concerne les points 2 et 3, c'est-à-dire, l'exploitation optimale du réseau sémantique réalisé pour la tâche de construction d'amorces de consensus.

La méthodologie a suivi deux étapes. Dans un premier temps, les mesures de similarité applicables à des réseaux sémantiques ont été identifiées puis adaptées au réseau sémantique construit. Dans un deuxième temps, un ensemble d'éditeurs d'ontologies a été étudié selon des critères issus de la littérature. Un éditeur a été choisi et des fonctionnalités lui ont été rajoutées. Finalement, cet environnement a été utilisé pour comparer les performances des différentes mesures de similarité implémentées et a permis d'opter pour une mesure particulière. La validité des amorces de consensus obtenues en pratique renseignera à terme sur la validité de l'organisation des concepts au sein du réseau sémantique.

2 Construction d'un réseau sémantique pour la description des images

L'idée du projet était de construire une ontologie des termes d'interprétation des images anatomopathologiques de cancer du sein et de formaliser cette ontologie dans l'objectif de construire une mesure de similarité entre termes permettant de comparer automatiquement les descriptions de différents experts utilisant ces termes. Bien que certaines classifications terminologiques médicales comportent une section anatomopathologique générale, comme la SNOMED-RT (Spackman *et al.*, 1997), l'UMLS (Lindberg *et al.*, 1993) ou GALEN (Rector *et al.*, 1995), aucune ontologie

n'était disponible formalisant les concepts de descriptions des images d'anatomopathologie mammaire pour un objectif spécifique de téléconsensus.

Les experts on souhaité que le modèle de description des images soit conforme aux données récentes de la littérature. Ils ont donc défini le vocabulaire contrôlé utilisable pour la description des images à partir de classifications diagnostiques et pronostiques publiées et validées (Consensus conference on the classification of ductal carcinoma in situ., 1997 ; Tavassolli & Deville, 2003). Les termes, organisés dans une taxinomie, correspondaient d'une part à des termes diagnostiques traduisant l'interprétation de lésions observées au niveau des images (taxinomie de diagnostics lésionnels) et d'autre part à des termes de description traduisant l'analyse sémiologique des images (taxinomie de caractéristiques morphologiques).

En ce qui concerne la taxinomie des termes de description, il est rapidement apparu que la relation taxinomique classique ne correspondait pas à une dimension de similarité intéressante ici. En effet, les caractéristiques décrites par les experts sont choisies pour leur valeur diagnostique discriminante. Par exemple, le mode de répartition des tubes est corrélé à la malignité des lésions. Dans une taxinomie, « tubes de répartition centrifuge » et « tubes de répartition organoïde » sont proches car ayant un père commun « tubes - mode de répartition » alors que dans la tâche de consensus, ces concepts sont incompatibles car évoquant des degrés de malignité différents. Le modèle de la tâche de consensus entre experts a permis d'identifier deux dimensions de similarité pertinentes – similarité morphologique et similarité diagnostique – et a fixé des contraintes pour l'organisation des termes (Steichen *et al.*, 2003).

Une relation de **voisinage morphologique** a été introduite pour « rapprocher » certains fils d'un même concept. Par exemple, un lien de voisinage morphologique est créé entre « petit noyau » et « noyau de taille moyenne », ainsi qu'entre « noyau de taille moyenne » et « grand noyau » mais pas entre « petit noyau » et « grand noyau ». De cette manière, la distance (en nombre d'arcs) entre « petit noyau » et « noyau de taille moyenne » (qui sont voisins morphologiques) est plus faible que celle entre « petit noyau » et « grand noyau ».

Une relation de **participation diagnostique** a été introduite pour rapprocher les termes complémentaires du point de vue diagnostique. Une hiérarchie diagnostique a été construite, permettant de rapprocher deux termes de description en les liant tous les deux au même diagnostic, par une relation « participe au diagnostic ». Ainsi, les termes de description « tubes de forme ramifiée » et « tubes de répartition centripète », cousins éloignés d'un point de vue taxinomique, deviennent voisins diagnostiques car ils pointent tous les deux vers le diagnostic « cicatrice radiaire ».

Finalement, le réseau sémantique créé comporte aujourd'hui une hiérarchie de 32 classes diagnostiques, un réseau de 177 classes morphologiques et deux types de liens supplémentaires autres que le lien taxinomique « is-a ».

3 Mesures de similarité dans les réseaux sémantiques

La similarité est une relation qui rend compte de certaines ressemblances entre des objets, des notions, des propriétés ou des relations. On distingue habituellement la similarité d'ensemble et la similarité dimensionnelle, qui concerne seulement quelques aspects limités de comparaison (Vosniadou & Ortony, 1989). L'intérêt de la similarité dimensionnelle est de permettre de former des catégories selon la dimension concernée. Les modèles de similarité reposent selon les cas (Bisson 2000) sur une quantification des propriétés (du contenu) partagées, ou sur une notion de proximité, variant en fonction inverse de la distance, dans un réseau.

3.1 Différents types de mesures dans les réseaux sémantiques

Les mesures de similarité dans les réseaux sémantiques sont classées selon l'information qu'elles utilisent (Budanitsky & Hirst, 1999) : les liens entre les termes, le contenu des termes, ou les deux.

L'approche par les liens assimile la distance entre deux termes au nombre de liens qui les séparent sur le plus court chemin dans le réseau (Leacock & Chodorow, 1998). Mais tous les liens dans un réseau ne correspondent pas à la même distance, et certains auteurs leurs attribuent un poids différent selon leur type, leur profondeur hiérarchique ou encore la densité locale du réseau (Sussna, 1993 ; Jiang & Conrath, 1997 ; Maynard & Ananiadou, 2001 ; McHale, 1998).

L'approche par le contenu suppose qu'on puisse en quantifier la différence entre deux termes. Le moyen le plus naturel est de comparer leurs descripteurs et attributs, mais elle nécessite une formalisation rigoureuse de la représentation des connaissances, qui est très coûteuse (Rodriguez, 2000). En pratique, c'est le contenu d'information des termes qui est utilisé (Resnik, 1995 ; Lin, 1998). Il dépend de la fréquence d'occurrence de leurs instances au sein d'un corpus : plus un terme est rare, et donc discriminant, plus son contenu d'information est élevé. Cette fois-ci, l'étape coûteuse est le recensement des occurrences au sein d'un corpus.

Le principe des mesures mixtes est de prendre pour distance le plus court chemin dans le réseau, et de pondérer les liens par la différence de contenu d'information entre le père et le fils (Richardson *et al.*, 1994 ; Jiang & Conrath, 1997). La similarité est ensuite calculée comme pour les mesures par les liens.

3.2 Adaptation des mesures de similarité au réseau sémantique d'IDEM

Dans le contexte du réseau sémantique d'IDEM, nous avons implémenté trois mesures : la mesure par les liens de (Leacock & Chodorow, 1998), figure 1, la mesure par le contenu de (Lin, 1998), figure 2, et la mesure mixte de (Jiang & Conrath, 1997) que nous avons adaptée pour prendre en compte les liens non taxinomiques.

Pour la mesure par le contenu, le calcul du contenu d'information est basé sur la fréquence d'occurrence dans un corpus composé de 710 comptes rendus produits par 10 pathologistes provenant de 2 institutions. Nous avons utilisé l'outil d'analyse

Syntex®[1] pour associer les termes du corpus et leurs instances avec les termes du réseau sémantique.

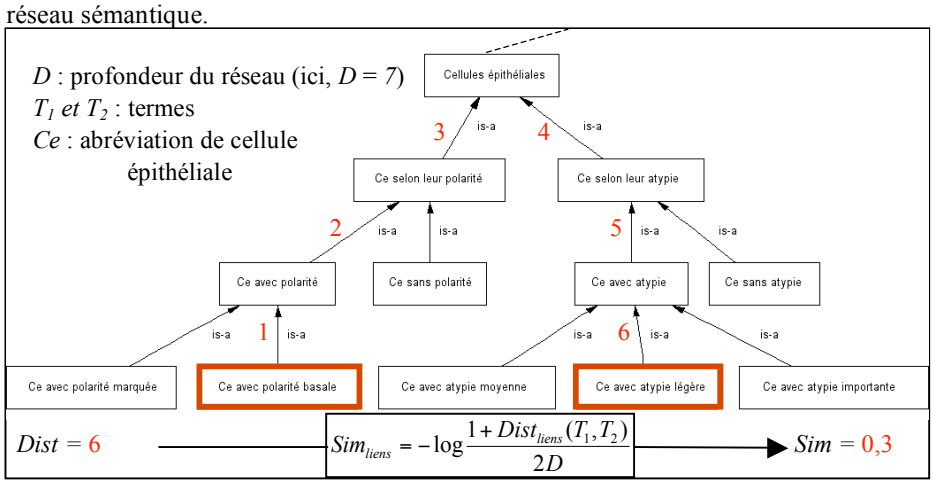

Fig. 1 – Mesure de similarité par les liens

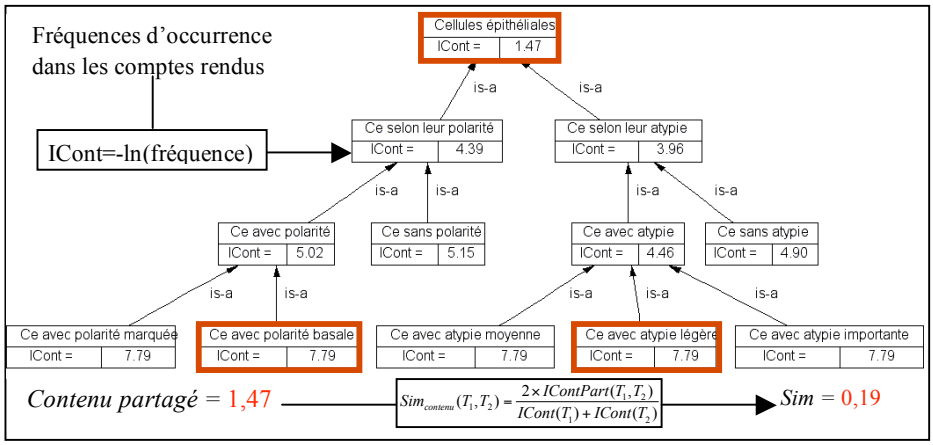

Fig. 2 – Mesure de similarité par le contenu

Dans sa forme originelle, la mesure mixte de Jiang prend en compte pour chaque lien la différence de contenu d'information du père et du fils [$CI(T_{fils})$ - $CI(T_{père})$]. Cette différence n'a de sens que pour les liens taxinomiques (la croissance du contenu d'information n'étant assurée par construction qu'en descendant la hiérarchie taxinomique). Pour les autres types de relations présentes au sein de notre réseau sémantique, nous avons remplacé cette différence par un coefficient $\Delta CI_{moy}(T)$ qui

[1] Syntex est développé par Didier Bourigault, ERSS, UMR 5610 CNRS, Toulouse

représente la moyenne des différences de contenus d'information pour les liens taxinomiques issus du même fils.

$$\Delta CI_{moy}(T) = \frac{\sum_{T_i \in Pères(T)} (CI(T) - CI(T_i))}{Nombre(Pères(T))} \quad (1)$$

4 Développement d'un environnement de gestion des connaissances

La gestion des connaissances et de leur exploitation pour déterminer des mesures de similarité a nécessité le développement d'un environnement informatique reposant sur un éditeur d'ontologie existant, mais offrant des fonctionnalités complémentaires.

4.1 Choix d'un éditeur d'ontologie

La taille des ontologies, la nature collaborative de leur construction, la nécessité de fréquemment les mettre à jour et l'intérêt potentiel de leur partage justifient l'emploi d'outils d'édition. Quelques revues récentes sur les éditeurs d'ontologie (Duineveld *et al.*, 2000 ; Gomez-Perez, 2002 ; Ribiere & Charlton, 2001) synthétisent les caractéristiques souhaitables pour ces outils dans cinq rubriques :

- Critères généraux concernant l'évaluation du logiciel (ergonomie, documentation, …),
- Critères liés à la structure ontologique (langage de représentation, contrôles de cohérence, …),
- Critères liés à la possibilité de coopération entre utilisateurs (travail synchrone et à distance, …),
- Critères liés à la simplicité de mise à jour de l'ontologie (modifications complexes, réversibilité, …),
- Critères liés aux possibilités d'interopérabilité entre systèmes.

Une fiche d'évaluation a été créée à partir de ces critères pour évaluer plusieurs éditeurs d'ontologie (Protégé, OilEd, …). Notre choix s'est porté sur Protégé 2000® (Noy *et al.*, 2000). Les concepts y sont représentés sous une forme générique, permettant de leur attribuer des propriétés et de les mettre en relation les uns avec les autres. Le choix de Protégé en tant qu'éditeur d'ontologie a été en partie motivé par sa capacité à accueillir des « plug-ins » qui étendent ses fonctionnalités.

4.2 Enrichissement de Protégé

La description d'attributs d'une ontologie sous Protégé se fait via des « Templates Slots », attributs d'un concept dont les valeurs sont héritées à tous ses fils. Dans le contexte du réseau sémantique d'IDEM, les attributs de l'ontologie doivent pouvoir

être valués directement (la fréquence d'occurrence, par exemple). Deux méta-classes ont été créées pour rendre possible l'implémentation du réseau sous Protégé :

- MainMetaClass : le formulaire de descriptions des concepts de l'ontologie
- SimilarityMetaClass : l'affichage de table des similarités entre termes

Les concepts de l'ontologie sont représentés comme instances d'une méta-classe et possèdent tous les attributs qui y sont définis. La figure 3 montre le formulaire de description des concepts.

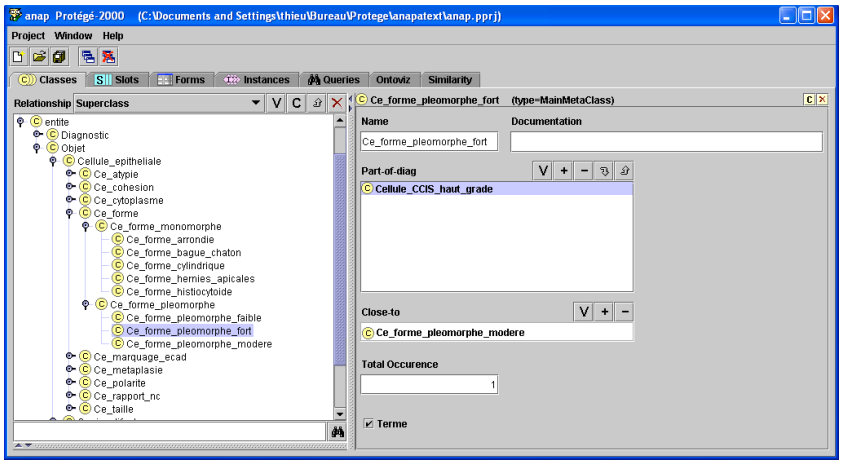

Fig. 3 – formulaire de description des concepts

Pour la comparaison automatique de deux concepts à l'aide de mesures de similarités, des « plug-ins » ont été développés avec les fonctionnalités suivantes :

- *Création automatique des méta-classes*

Les méta-classes sont créées dynamiquement à chaque démarrage de Protégé. L'avantage de la construction automatique de ces méta-classes est l'assurance que l'ontologie sera toujours construite et toujours mise à jour suivant la même logique du formulaire de description.

- *Saisie des attributs*

Pour faciliter la construction de l'ontologie, un widget (plug-in de saisie) particulier a été réalisé pour l'attribut « participe au diagnostic ». En effet, la création d'un lien de ce type peut entraîner le besoin de créer des liens similaires hérités. Par exemple, si on ajoute au concept A un attribut « participe au diagnostic » pointant vers le diagnostic B, plusieurs situations sont envisageables (figure 4) :

le concept A a des liens « participe au diagnostic » pointant vers tous les fils de B.

tous les fils de A ont un lien « participe au diagnostic » pointant vers B.

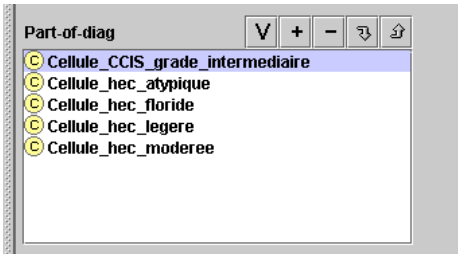

Fig. 4 – plug-in de saisie pour l'attribut « participe au diagnostic »

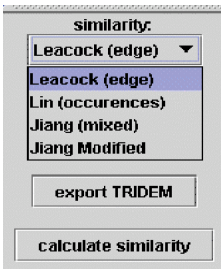

Fig. 5 – méthodes de calcul de similarités

- *Calcul de similarité*

L'extension de Protégé vise à fournir un calcul de similarité entre concepts de l'ontologie. C'est pourquoi un plug-in a été développé permettant le choix entre plusieurs méthodes de calcul de similarités qui ont été implémentées (figure 5).

- *Affichage de la table des similarités*

Un plug-in d'affichage a été créé pour présenter les similarités entre tous les concepts 2 à 2. Les similarités supérieures à un seuil donné (0,6 est choisi ici arbitrairement) sont colorées (figure 6). En théorie, les experts devraient parcourir cette table et valider chaque valeur de similarité, ce qui est une tâche difficile et c'est pourquoi d'autres outils graphiques ont été réalisés.

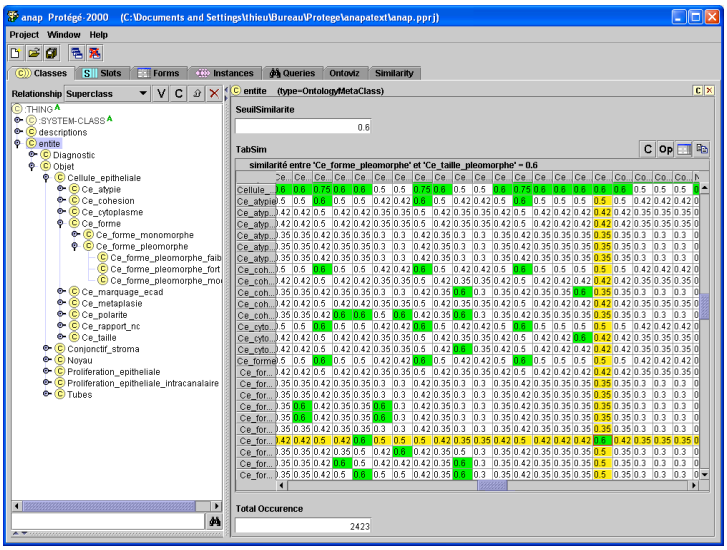

Fig. 6 – plug-in d'affichage de la table des similarités

- *Affichage des similarités*

Un autre « plug-in » d'affichage des similarités a été développé, dont l'objectif est la validation des résultats de similarité par l'expert. Il permet de visualiser clairement la similarité entre descriptions complexes (qui contiennent plusieurs termes descriptifs) ou bien directement entre termes descriptifs, en affichant graphiquement le plus court chemin entre les deux termes considérés (figure 7).

- *Visualisation graphique des concepts similaires dans l'ontologie*

Un « plug-in » existant OntoViz, libre (open source), permet la représentation graphique d'ontologie. Il a été réalisé par Michael Sintek (à Stanford Medical Informatics) et utilise la librairie libre GraphViz en Java (fournie par AT&T) qui permet de visualiser des graphes.

Ce plug-in a été amélioré pour pouvoir naviguer plus facilement dans l'ontologie et afficher, en couleur, tous les termes similaires à un terme donné.

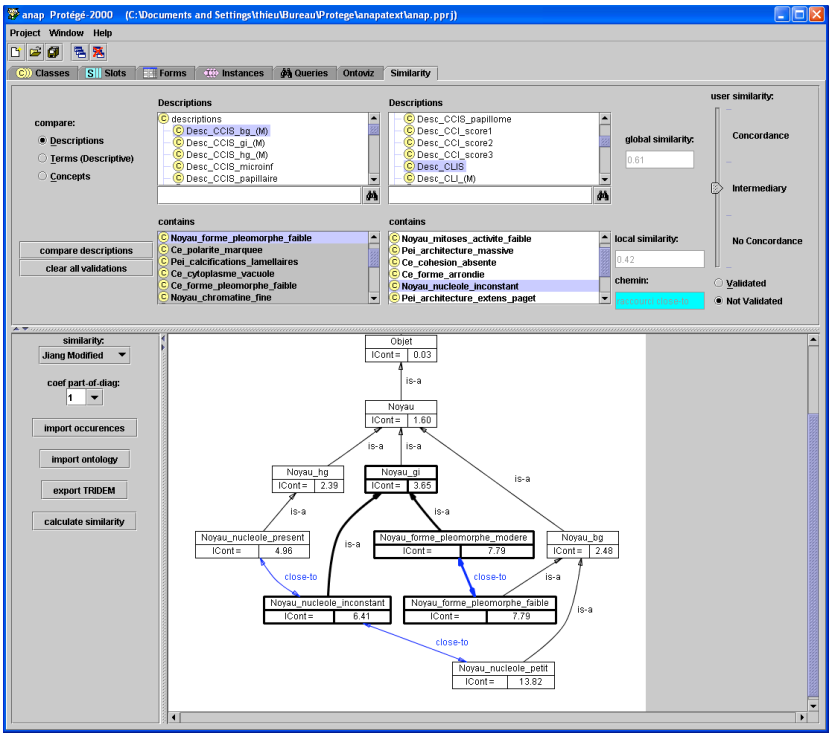

Fig. 7 – validation des similarités

5 Comparaison des résultats de similarité

Chaque résultat de similarité dépend à la fois de la mesure utilisée et du réseau sémantique sur lequel la mesure est appliquée. Dans l'évaluation menée, plusieurs mesures sont appliquées sur le même réseau de sorte à déterminer le rôle joué par le

choix d'une mesure de similarité particulière. Les résultats de similarité donnés par les différentes mesures ont été exportés et analysés au moyen de Stata®[2].

Dans un premier temps, la variabilité entre mesures a été évaluée par le coefficient de corrélation intraclasse (CCI) de Hoyt (Falissard, 2001), indicateur de variabilité pour les mesures continues, d'autant plus proche de 1 que cette variabilité est faible et donc que les mesures sont cohérentes. La variabilité entre mesures prises deux à deux est importante, avec des CCI qui vont de 0,61 [0,59–0,62] entre Jiang et Leacock, à 0,72 [0,71–0,73] entre Jiang et Lin. Cela signifie que les résultats de similarité donnés par les différentes mesures sont différents. En analysant ces différences, on observe que lorsque les termes sont très proches ou très éloignés dans le réseau, les diverses mesures s'accordent à les reconnaître respectivement fortement ou faiblement similaires. Par contre, lorsque les termes sont en position intermédiaire dans le réseau, les mesures donnent des résultats nettement divergents.

Dans un deuxième temps, nous souhaitions déterminer si les mesures de similarité calculées entre termes du réseau donnaient des résultats correspondants à ceux qu'attendent des utilisateurs exigeants dans le contexte de la tâche que le système doit assister. Trois experts ont émis des jugements de similarité, allant de « absolument concordants » à « absolument discordants » et comportant deux graduations intermédiaires « plutôt discordants » et « plutôt concordants », sur un échantillon aléatoire de 200 couples de termes de description. Les variabilités intra- et inter-experts des jugements ont été analysées par coefficient kappa de Cohen (Falissard, 2001), d'autant plus proche de 1 que cette variabilité est faible et donc que les experts sont convergents. La variabilité intra- et inter- observateurs des trois experts sur l'échantillon de 200 couples est importante (kappa de l'ordre de 0,50).

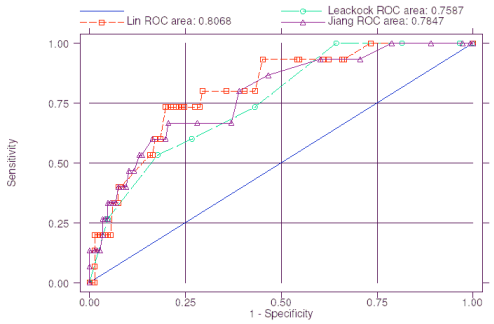

Fig. 8 – Courbes ROC

Un étalon or a donc été construit en classant « similaires » les couples que les trois experts jugeaient unanimement « absolument concordants » ou « plutôt concordants », et « dissimilaires » les autres. La valeur discriminante des mesures a alors pu être évaluée en traçant leurs courbes ROC (« Receiver Operating Characteristics »), un moyen d'exprimer la relation entre la sensibilité et la spécificité

[2] Stata est un logiciel scientifique de statistique distribué par la société « Integral Software », Paris (http://www.stata.com/)

des mesures envisagées (figure 8). Dans tous les cas, l'aire sous la courbe est bonne (entre 0,76 [0,64-0,87] pour Leacock et 0,81 [0,70-0,91] pour Lin), laissant préjuger de propositions d'amorces de consensus utiles (la bonne sensibilité des méthodes permet de rapprocher des situations que l'expert aurait lui-même rapprochées). La différence entre les performances des trois mesures n'est pas statistiquement significative.

6 Discussion et Conclusion

L'environnement construit à partir de Protégé permet une construction facilitée de l'ontologie et une comparaison efficace des mesures de similarités. Il assure également une gestion contrôlée de la pertinence pour le problème posé des mesures de similarité lorsque le réseau sémantique évolue.

Les résultats de l'évaluation montrent que chacune des trois mesures peut produire des valeurs de similarité utiles pour aider les pathologistes à atteindre le consensus. D'autres arguments ont été considérés pour effectuer un choix final. Le premier argument est l'expressivité de la mesure. Il est important que les experts appréhendent la signification des résultats de similarité obtenus pour n'importe quelle paire de termes. Ce critère est difficilement rempli par la mesure par le contenu car ce contenu n'est pas intuitif. Le deuxième argument est le coût élevé du calcul de contenu d'information, en particulier du recensement des occurrences, ce qui favorise aussi l'utilisation de mesures par les liens.

En résumé, nous avons mis en œuvre avec succès trois mesures de similarité dans le réseau sémantique de l'application, ainsi que des fonctionnalités de validation. Les mesures ont été comparées au jugement de similarité des experts et la mesure par les liens est aussi performante que les mesures par le contenu. Cela contredit des résultats sur d'autres réseaux sémantiques et suggère que l'utilisation de liaisons non-taxinomiques dans notre réseau permet d'éviter le besoin de calculer le contenu d'information. En absence de différence statistiquement significative entre les résultats des différentes mesures, nous avons employé des arguments pragmatiques pour choisir la mesure par les liens, plus lisible et moins coûteuse.

Une amélioration future concerne la correction du réseau sémantique. En effet, il est utopique de penser aboutir à un réseau figé et parfait. Dès lors, des corrections devront être appliquées au réseau pour obtenir de meilleurs résultats de similarité. Il faudrait alors implémenter des fonctionnalités qui permettent le (re)positionnement automatique d'un terme en fonction de contraintes de similarité avec les autres termes.

Références

BISSON G. (2000) La similarité : une notion symbolique/numérique. Diday E, Kodratoff Y., Brito P., Moulet M. (eds) Induction symbolique/numérique à partir de données, Cepadues

BUDANITSKY A., HIRST G. (2001). « Semantic Distance in WordNet: An Experimental, Application-oriented Evaluation of Five Measures », *Workshop on WordNet and Other Lexical Resources, in the North American Chapter of the Association for Computational Linguistics*, Pittsburgh, 2001.

Consensus conference on the classification of ductal carcinoma in situ. Hum Pathol. 1997 Nov; 28(11): 1221-5.

DE VET H.C.W., KOUDSTAAL J., KWEE W.-S., WILLEBRAND D., ARENDS J.W. (1995). « Efforts to Improve Interobserver Agreement in Histopathological Grading », *Journal of Clinical Epidemiology*, vol. 48, n° 7, 1995, p. 869-73.

DUINEVELD A., STOTER R., WEIDEN M., KENEPA B., BENJAMINS V. (2000). « Wondertools ? A Comparative Study of Ontological Engineering Tools », *International Journal of Human-Computer Studies*, vol. 52, n° 6, 2000, p. 1111-33.

FALISSARD B. (2001). Mesurer la subjectivité en santé: perspective méthodologique et statistique, Paris, Masson, 2001.

FLEMING K.A. (1966). « Evidence-based pathology », J Pathol, 1996 Jun; 179(2): 127-8

GOMEZ-PEREZ A. et al. (2002). Ontology building tools, Gomez-Perez A., A survey on ontology tools, *OntoWeb Consortium*, 2002: 13-34.

JIANG J., CONRATH D. (1997). « Semantic Similarity Based on Corpus Statistics and Lexical Taxonomy », *Proceedings of the 10th International Conference on Research on Computational Linguistic*, Taïwan, 1997.

LEACOCK C., CHODOROW M. (1998). « Combining Local Context and WordNet Similarity for Word Sense Identification », in Felbaum C., *WordNet: An Electronic Lexical Database*, Cambridge, MA, The MIT Press, 1998, p. 265-283.

LE BOZEC C., JAULENT M.-C., ZAPLETAL E. (2000). « IDEM : remémoration de cas pour l'aide au diagnostic en anatomie pathologique », in Charlet J., Zacklad M., Kassel G., Bourigault D., *Ingénierie des connaissances, évolutions récentes et nouveaux défis*, Paris, Eyrolles, 2000, p. 371-386.

LIN D. (1998). « An Information-Theoretic Definition of Similarity », *International Conference on Machine Learning*, 1998, p. 296-304.

LINDBERG D.A., Humphreys B.L., McCray A.T. (1993). « The Unified Medical Language System », *Methods Inf Med*, vol. 32, n° 4, 1993, p. 281-91.

MCHALE M. (1998). A comparison of WordNet and Roget's taxonomy for measuring semantic similarity. Harabagiu S. *Proceedings of the COLING/ACL 1998 Workshop on Usage of WordNet in Natural Language Language Processing Systems*. Montréal: Presses de l'université de Montréal, 1998: 115-20.

MAYNARD D., ANANIADOU S. (2001). « Term Extraction Using a Similarity-Based Approach », in Bourigault D., Jacquemin C., L'homme M., *Recent Advances in Computational Terminology*, Amsterdam, John Benjamins, 2001, p. 261-78.

NOY N., FERGERSON R., MUSEN, M. (2000). « The Knowledge Model of Protege-2000: Combining Interoperability and Flexibility », *2nd International Conference on Knowledge Engineering and Knowledge Management (EKAW'2000)*, Juan-les-Pins, 2000.

RECTOR A.L., SOLOMON W.D., NOWLAN W.A., RUSH T.W., ZANSTRA P.E., CLAASSEN W.M. (1995). « A Terminology Server for Medical Language and Medical Information Systems », *Methods Inf Med*, vol. 34, n° 1-2, 1995, p. 147-57.

RESNIK P. (1995). Using information content to evaluate semantic similarity in a taxonomy. *Proceedings of the 14th International Joint Conference on Artificial Intelligence*. 1995.

RIBIERE M., CHARLTON P. (2001). Ontology overview. *Motorola labs*, 2001.

RICHARDSON R., SMEATON A., MURPHY J. (1994). Using WordNet as a knowledge base for measuring semantic similarity between words. *Proceedings of AICS Conference*. 1994.

RODRIGUEZ, MA. (2000). Assessing Semantic Similarity among Spatial Entity Classes, *PhD thesis*, University of Maine, 2000.

STEICHEN O., LE BOZEC C., THIEU M., ZAPLETAL E., JAULENT M.-C. (2003). Construction d'un réseau sémantique supportant des mesures de similarité pour une aide au consensus en imagerie médicale, *Actes du colloque CITE 2003*, Troyes, pp. 93-110.

SUSSNA M. (1993). « Word Sense Disambiguation for Free-text Indexing Using a Massive Semantic Network », *Second International Conference on Information Knowledge Management, CIKM'93*, Arlington, 1993, p. 67-74.

TAVASSOLI F., DEVILLE P. (2003). Pathology and Genetics of Tumours of the Breast and Female Genital Organs; *SeriesIARC/World Health Organization Classification of Tumours*, 2003.

VOSNIADOU S., ORTONY A. (1989). « Similarity and Analogical Reasonning: a Synthesis », in Vosniadou S., Ortony A., *Similarity and Analogical Reasoning*, Cambridge, Cambridge University Press, 1989, p. 1-17.

ZAPLETAL E., LE BOZEC C., DEGOULET P., JAULENT MC. (2003). « A collaborative platform for consensus sessions in pathology over Internet », Stud Health Technol Inform. 2003;95:224-9.