

# Système d'aide à la décision pour la surveillance de la qualité de l'air intérieur

Zoulikha Heddadji\*\*\*, Nicole Vincent\*, Severine Kirchner\*\*, Georges Stamon\*

\*Université René Descartes-Crip5-Équipe SIP-UFR Mathématiques et Informatique  
45, rue des Saints Pères 75006 Paris

{nicole.vincent,georges.stamon}@math-info.univ-paris5.fr

\*\*CSTB-84, avenue Jean Jaurès Champs-sur-Marne 77421 Marne-La-Vallée Cedex2  
zoulikha.heddadji@cstb.fr, kirchner@cstb.fr

## 1 Contexte

L'ensemble des mandats d'intervention en lien avec une mauvaise qualité de l'air dans les espaces fermés et leurs différentes solutions constitue notre base documentaire qui symbolise l'image représentative de la connaissance et du savoir faire des experts. Le raisonnement que nous mobilisons autour de cette base pour notre système d'aide à la décision est le raisonnement à partir de cas. L'interface interactive que nous avons développée nous permet de sauvegarder l'ensemble des cas dans un formalisme XML reproduisant la structure logique des plaintes. Les modèles de balises correspondent aux différentes rubriques citées dans les textes des plaintes (antécédents, symptômes, environnement, etc.). Le RàPC s'appuie d'abord sur le module de remémoration des cas de pollution similaires au contexte de la plainte courante. Ce module est fortement lié au contenu et à la structure des cas situés en mémoire. Le RàPC repose aussi sur le module d'adaptation qui ajuste le rapport du cas jugé le plus similaire au contexte du cas courant. Il est constitué également du module de consensus assurant la validation humaine multi-experte des cas, et du module d'apprentissage du nouveau cas pour capitaliser la nouvelle expérience en mémoire.

## 2 Phase de remémoration

Nous nous basons sur la mesure de similarité conceptuelle de (Zarga et Salotti, 2004) pour établir ensuite une distance sémantique entre le problème énoncé dans la plainte cible et les problèmes sources. Inspirées des travaux de (Wu et Palmer, 1994) Zarga et Salotti utilisent une mesure de spécificité qui favorise les liens père-fils par rapport aux autres liens :

$$Spec(C_1, C_2) = Depth_{bottom}(PPS(C_1, C_2)) \frac{Dis(PPS(C_1, C_2), C_1) Dis(PPS(C_1, C_2), C_2)}{2 \text{ depth } PPS_{C_1, C_2}}$$
$$Sim_{ZS}(C_1, C_2) = \frac{2 \text{ depth } PPS_{C_1, C_2}}{\text{depth}_{PPS_{C_1, C_2}}(C_1) + \text{depth}_{PPS_{C_1, C_2}}(C_2) + Spec(C_1, C_2)}$$

où  $Depth_{bottom}(PPS(C_1, C_2))$  est le nombre maximum d'arcs séparant le concept *bottom* ( le concept le plus bas du réseau conceptuel) du plus petit subsumant de  $C_1$  et  $C_2$ ;  $PPS(C_1, C_2)$ .  $Dis(C_1, C_2)$  est la distance en nombre d'arcs entre  $C_1$  et  $C_2$ .

## 2.1 Modèle de proximité flou

(Mercier & Beigbeder, 2004) estiment que plus les termes de la requête apparaissent proches au niveau d'un élément de la base plus ce dernier est pertinent par rapport à la requête cible. Le modèle de proximité flou du terme A par rapport au terme B est formalisé par:

$$\mu_{NEAR(A,B)}(d) = \text{Max}_{\substack{i \in d^{-1}(A) \\ j \in d^{-1}(B)}} \left( \text{Max} \left( \frac{k - |j - i|}{k}, 0 \right) \right)$$

Où  $d^{-1}(t)$  désigne l'ensemble des positions prises par le terme  $t$ , et  $k$  une constante fixe choisie représentant la taille de la fenêtre glissante des cooccurrences des termes.

## 2.2 Notre nouveau modèle: le modèle de proximité flou sémantique

La mesure de (Mercier & Beigbeder, 2004) est très intéressante, néanmoins elle ne tient pas compte de la sémantique des termes (dans le cas où des termes sémantiquement proches des termes utilisés dans la requête apparaissent directement proches au niveau d'un élément de la base). En effet, ce modèle est limité par la relation de cooccurrence directe des termes qui ne permet pas de capturer la proximité sémantique entre les mots. L'équation présentée dans le modèle de Mercier et Beigbeder devient :

$$\mu_{NEAR(A,B)}(d) = \text{Max}_{\substack{i \in d^{-1}(Simens(A)) \\ j \in d^{-1}(Simens(B))}} \left( \text{Max} \left( \frac{(k - |j - i|) Sim_{ZS}(ti, A) Sim_{ZS}(tj, B)}{k}, 0 \right) \right)$$

*Simens(A)* est l'ensemble des termes proches de A suivant la mesure d'appariement conceptuelle utilisée. Notre nouveau modèle apporte la connaissance de la sémantique au modèle existant. Les résultats des degrés de proximité que nous avons obtenus à l'étape d'expérimentation sur des données réelles et sur des données simulées sont très améliorés à l'aide de l'augmentation sémantique.

## Références

- Zargayouna, H et S. Salotti (2004). « Mesure de similarité dans une ontologie pour l'indexation sémantique de documents XML ». *IC'2004*.
- Wu, Z. et M. Palmer (1994). « Verb Semantics and Lexical Selection ». *Proceedings of the 32nd Annual Meetings of the Associations for Computational Linguistics*, 133-138.
- Beigbeder, M. et A. Mercier (2004). « Fuzzy set theory applied to a proximity model for information retrieval ». Nantes, France. LFA, 231-237.

## Summary

We present a new information retrieval model based on the semantic proximity level of term occurrences by using a conceptual network to identify the closest cases of the used request. Using this framework, we aim to implement a case based reasoning approach to help decision making in situations where indoor air is suspected to be responsible of health effects.