# Modèle de recherche d'information sémantique en graphe : interrogation par propagation d'activation

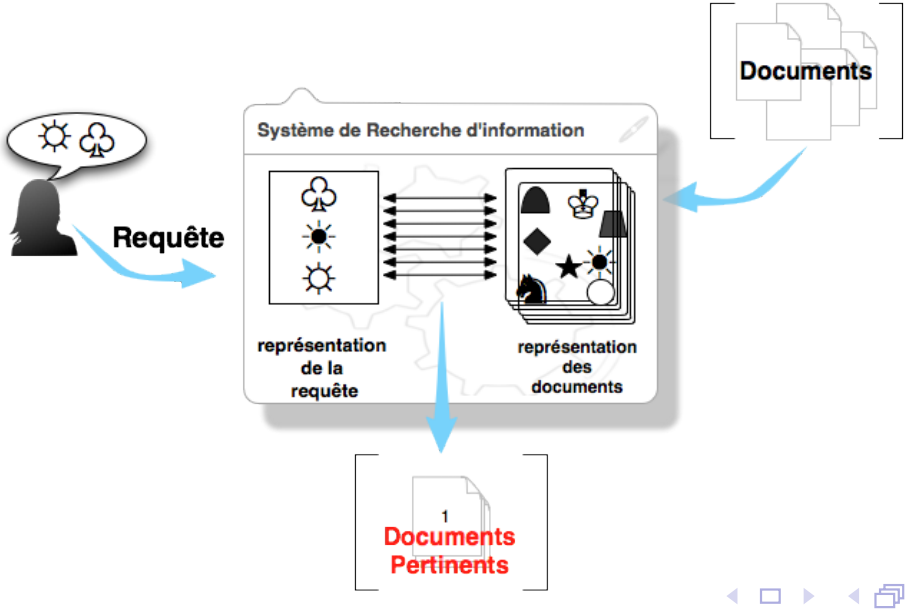
Ines Bannour, Haïfa Zargayouna, Adeline Nazarenko

LIPN, UMR 7030 CNRS, Université Paris 13, Sorbonne Paris Cité

6 juillet 2017

# Contexte : recherche d'information

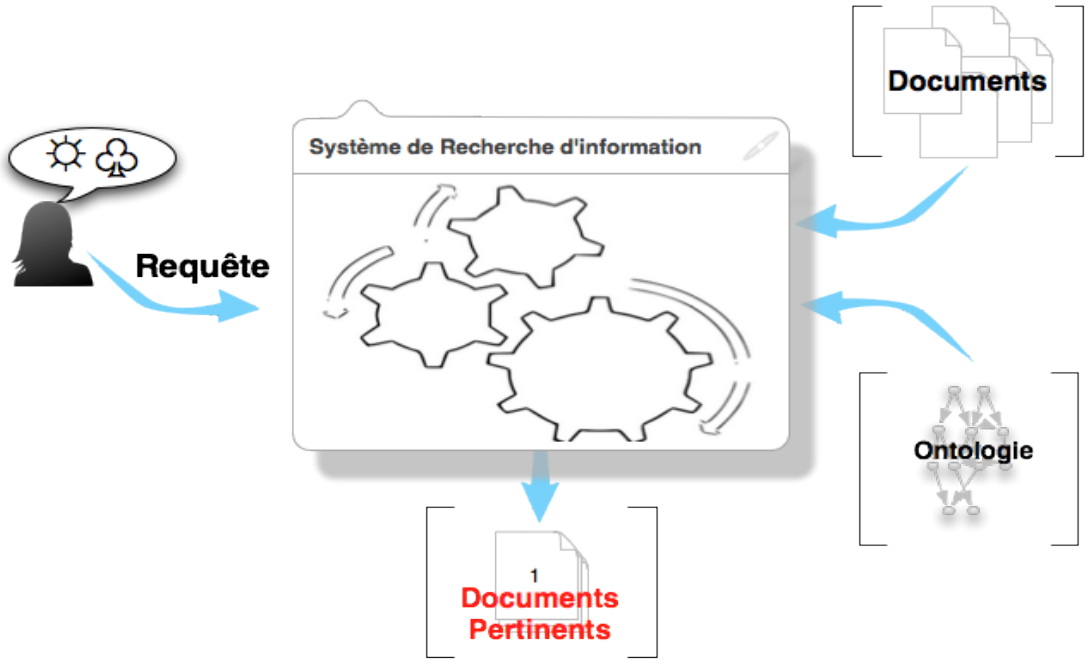
- La recherche d'information (RI) permet de caractériser et de retrouver un ensemble de documents pertinents par rapport à un besoin en information de l'utilisateur
- Un modèle de RI permet de choisir et définir
  - une représentation des documents (et des requêtes)
  - une fonction de correspondance entre la représentation de chaque document et la requête pour calculer le degré de similarité



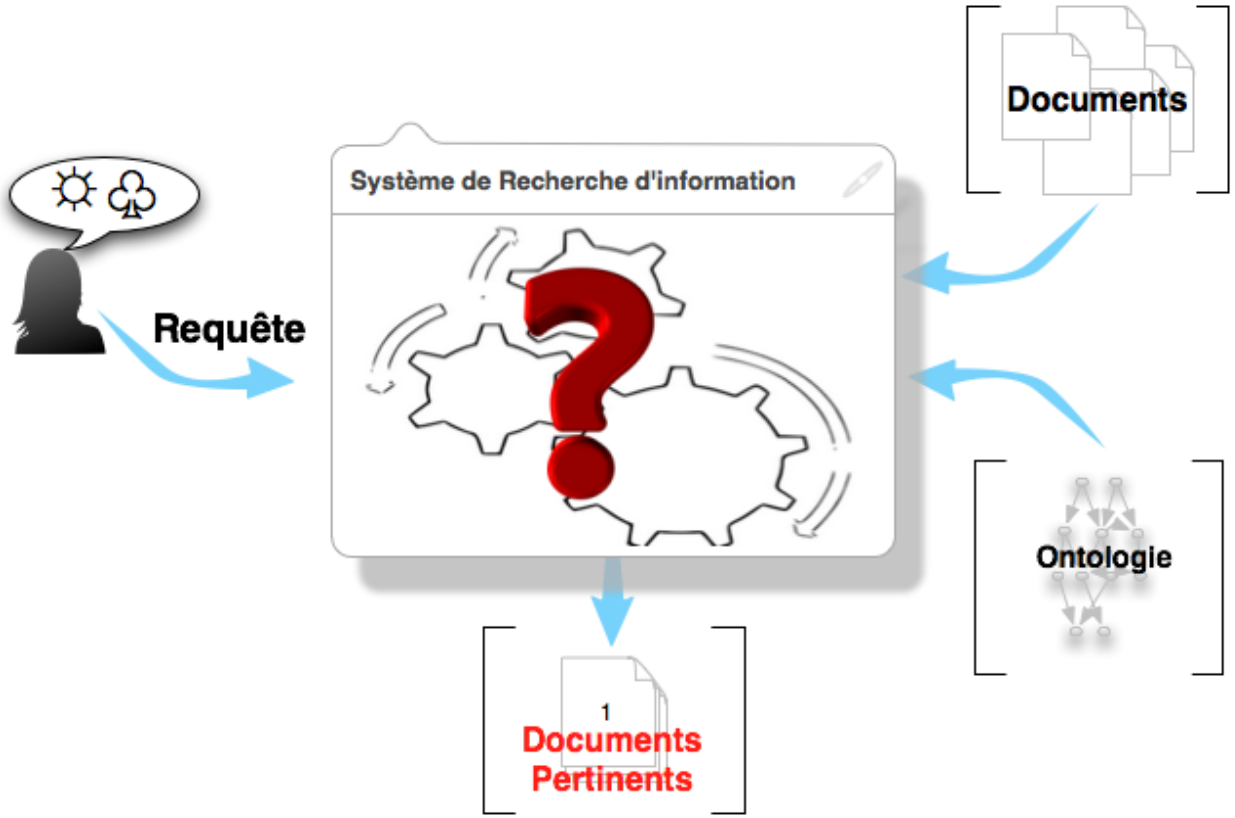
# Contexte : recherche d'information sémantique

La RI sémantique (RIS) :

- prendre en compte la richesse terminologique (synonymie, polysémie) en s'appuyant sur des ressources sémantiques externes



# Positionnement



# Plan


1. État de l'art
2. Modèle : graphe et propagation d'activation
3. Expérimentations
4. Conclusion et perspectives

# RIS et WS

- Modèles de RIS : adaptations des modèles classiques
  - ✓ Analyses distributionnelles classiques
  - ✗ Indépendance entre les termes, les documents et les concepts

  ⇒ Modèles sémantiques aplatis et appauvris

- Accès à l'information dans le WS
  - ✗ Recherche booléenne exacte
  - ✗ Contenu textuel du document non exploité
  - ✓ Traitement des requêtes complexes + exploitation de la richesse des ontologies :
    - concepts
    - relations hiérarchiques
    - relations du domaine
    - axiomes

# Hybridation RI et WS

- Approches hybrides "boite noire"
  - Hybridation des requêtes : mots clés + requêtes SPARQL
  - Hybridation des résultats : filtrage [Fernandez et al., 2011] + fusion ou intersection [Bhagdev et al., 2008, Bikakis et al., 2010]
    $\Rightarrow$ Agrégation des résultats
- Approches hybrides intégrées : modèles unifiés
  - Modèles de RI adaptés : Semplore [Zhang et al., 2007]
    $\Rightarrow$ Pas de raisonnement sur les ontologies
  - Modèles du WS adaptés : $CE^2$ [Wang et al., 2011]
    $\Rightarrow$ Pas de calcul distributionnel dans l'appariement

# Motivations : une véritable combinaison des modèles documentaires et sémantiques

- Modélisation : intégrer dans une même représentation
  - modèle documentaire
  - modèle sémantique

  en prenant en compte des données hétérogènes textuelles et sémantiques

# Motivations : une véritable combinaison des modèles documentaires et sémantiques

- Modélisation : intégrer dans une même représentation
  - modèle documentaire
  - modèle sémantique

  en prenant en compte des données hétérogènes textuelles et sémantiques

  ⇒ Un modèle en graphe sémantico-documentaire

# Motivations : une véritable combinaison des modèles documentaires et sémantiques

- Modélisation : intégrer dans une même représentation
  - modèle documentaire
  - modèle sémantique

  en prenant en compte des données hétérogènes textuelles et sémantiques

  ⇒ Un modèle en graphe sémantico-documentaire
- Interrogation : exploiter la complémentarité entre
  - analyses distributionnelles des approches numériques
  - analyses sémantiques des approches symboliques

# Motivations : une véritable combinaison des modèles documentaires et sémantiques

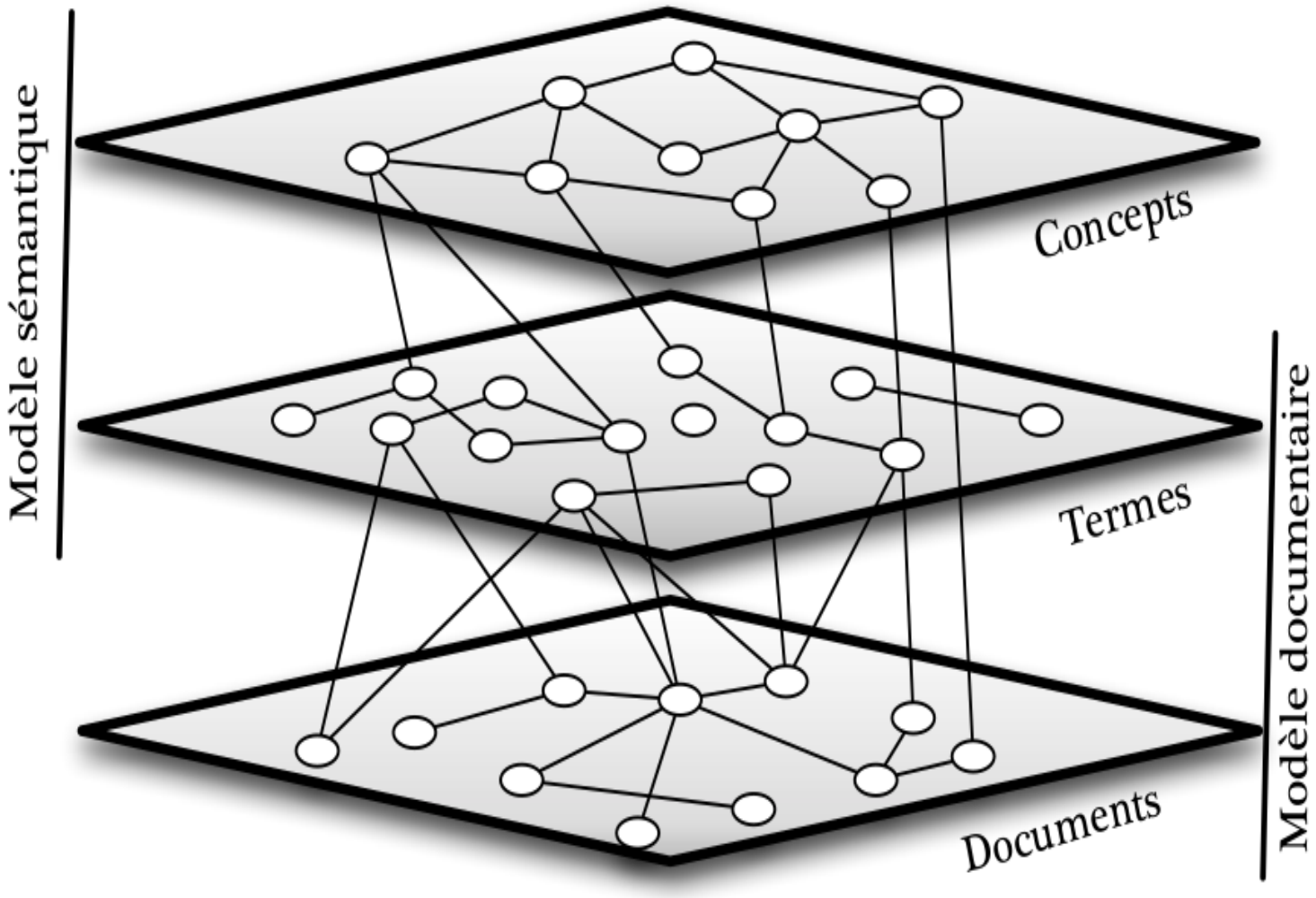
- Modélisation : intégrer dans une même représentation
  - modèle documentaire
  - modèle sémantique

  en prenant en compte des données hétérogènes textuelles et sémantiques

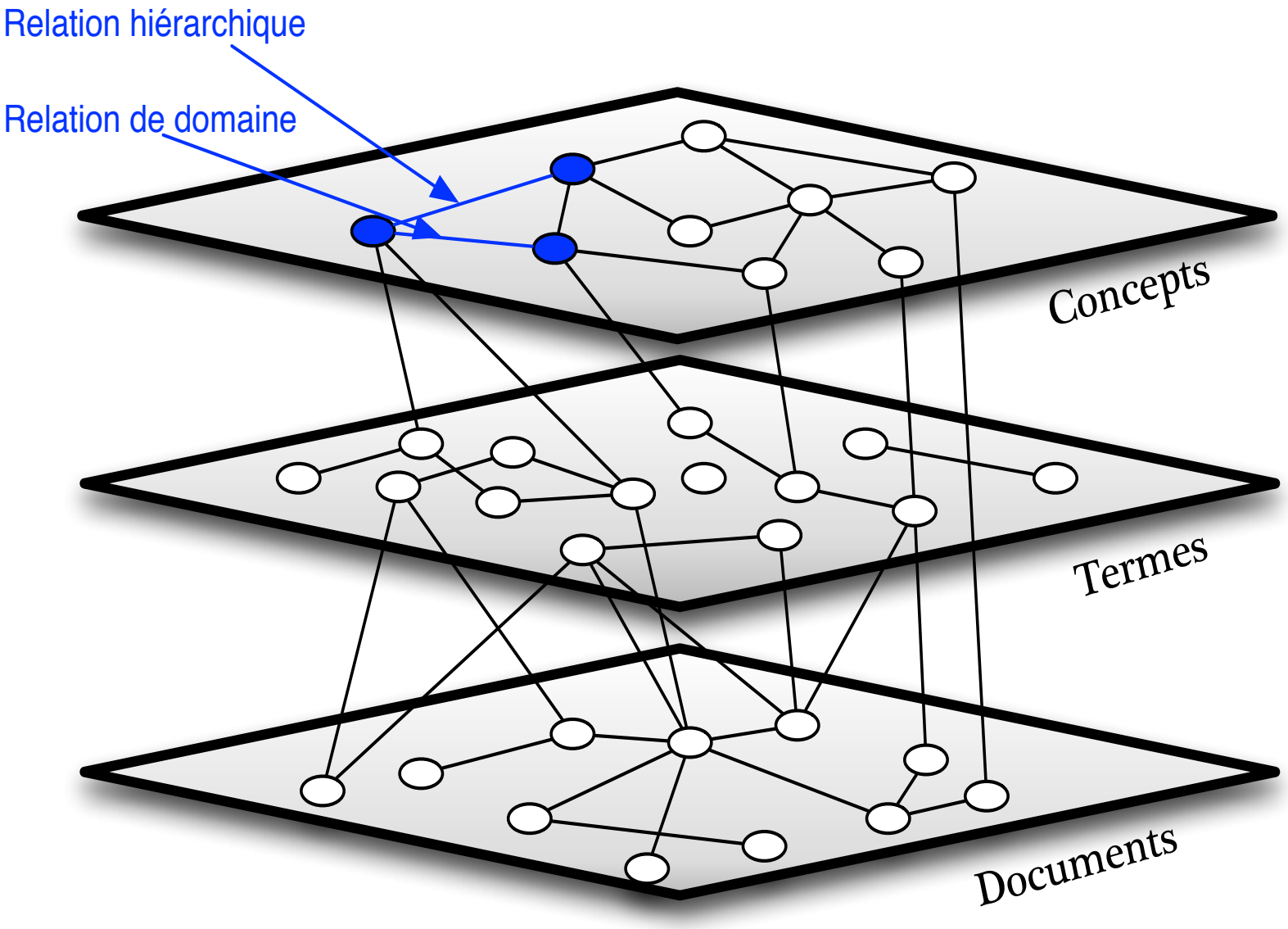
  ⇒ Un modèle en graphe sémantico-documentaire
- Interrogation : exploiter la complémentarité entre
  - analyses distributionnelles des approches numériques
  - analyses sémantiques des approches symboliques

  ⇒ Re-visiter l'algorithme de propagation d'activation

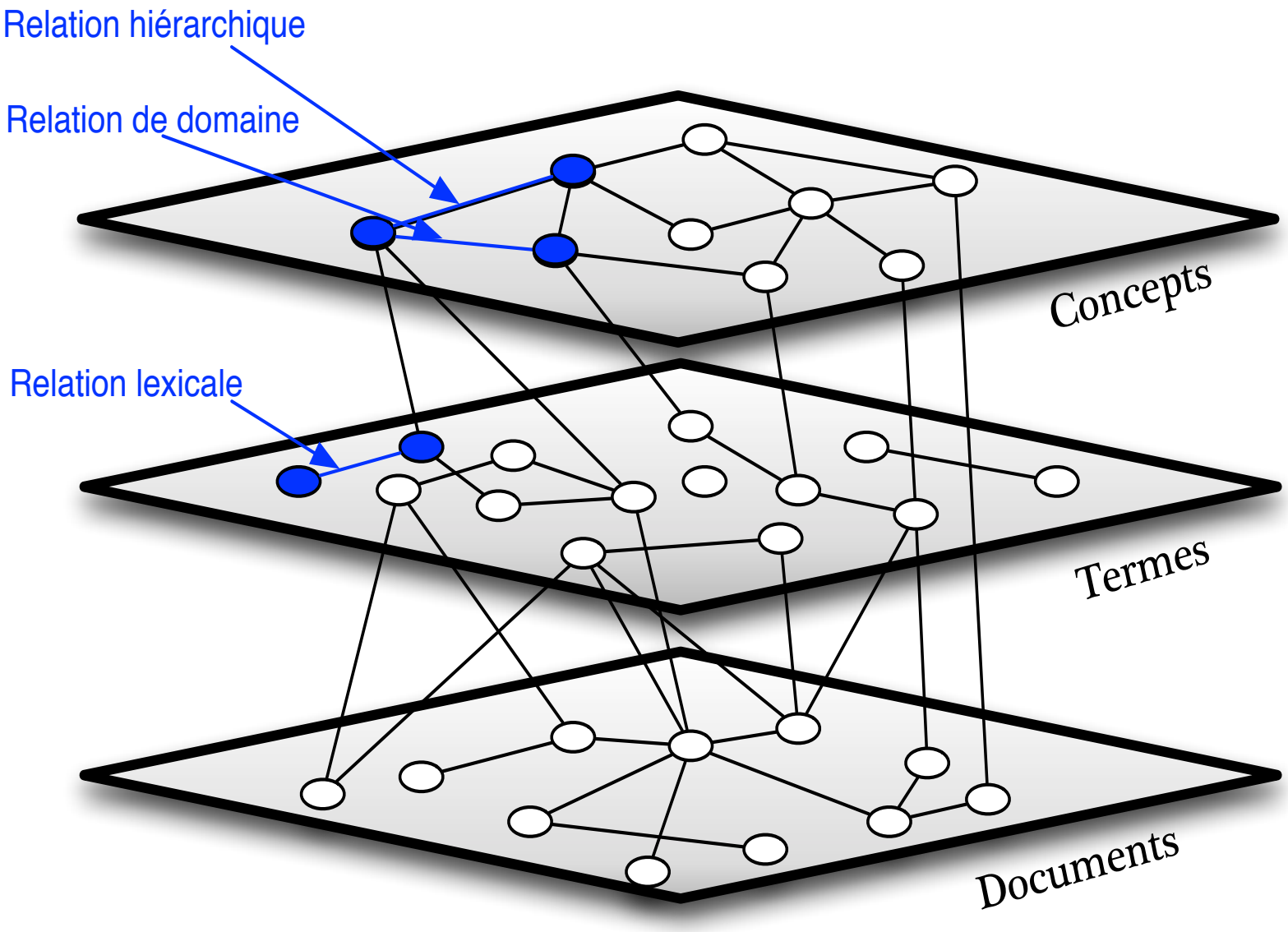
# Réseau sémantico-documentaire : 3 niveaux



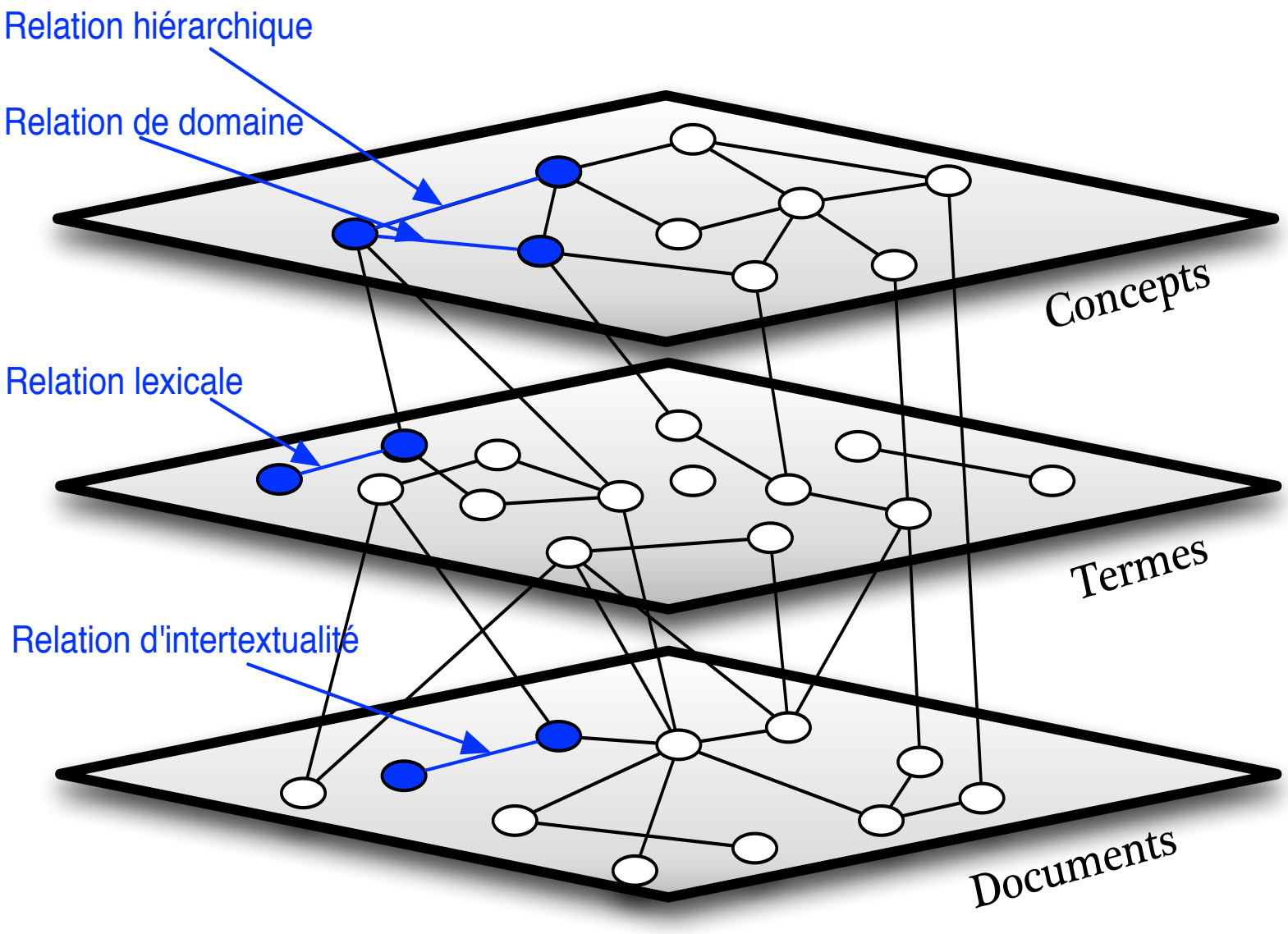
# Réseau sémantico-documentaire : nœuds et relations



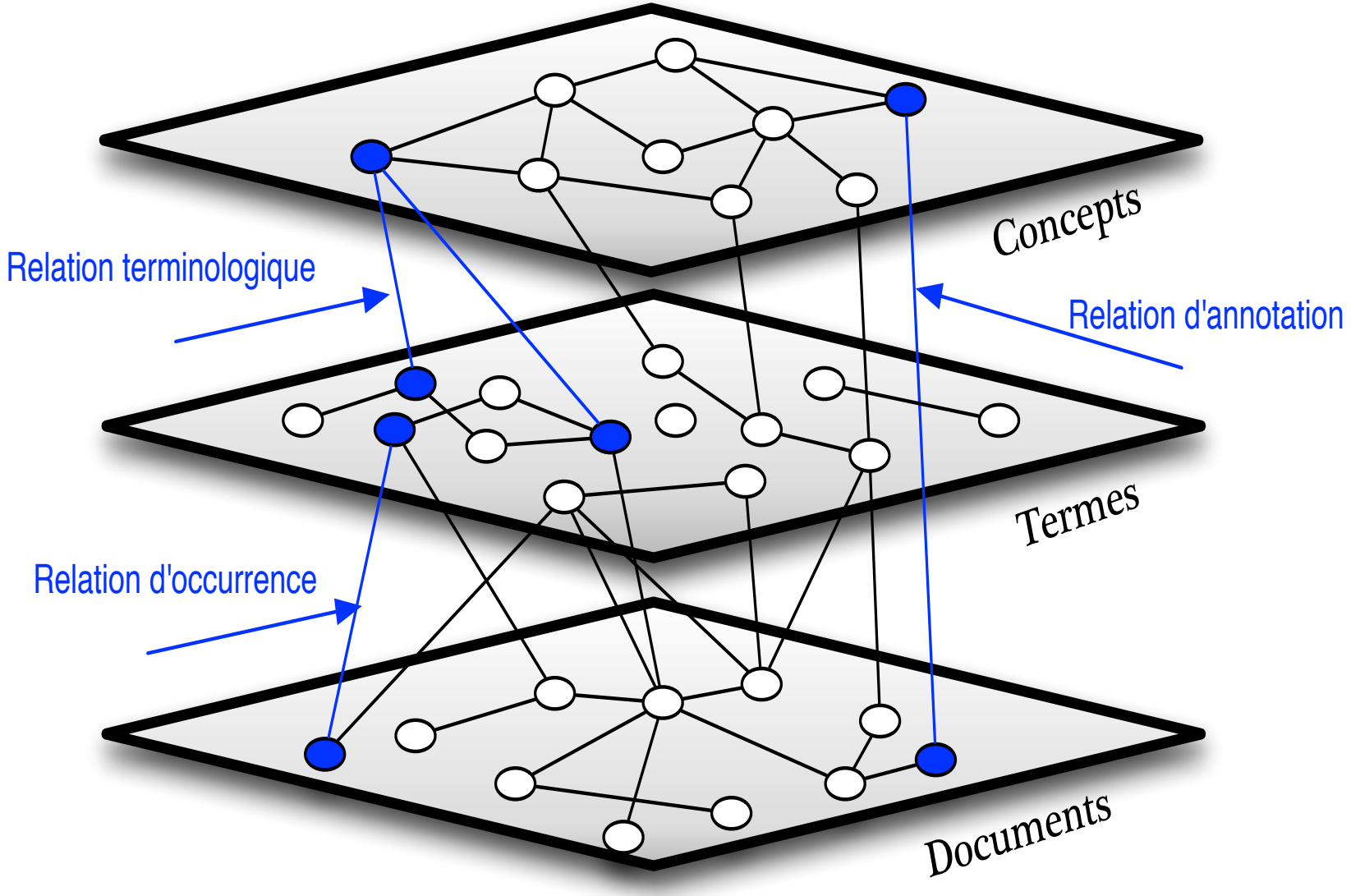
# Réseau sémantico-documentaire : nœuds et relations



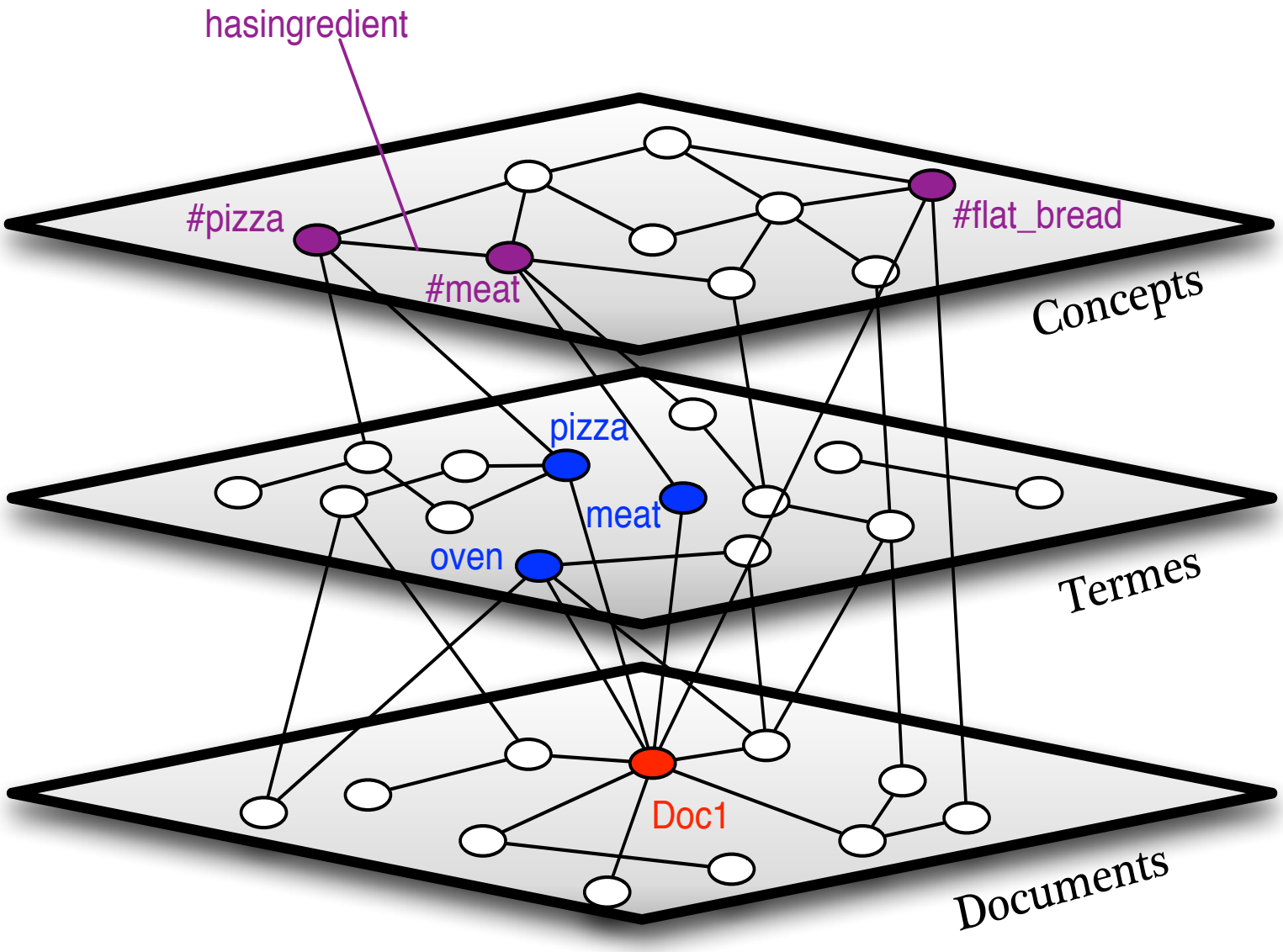
# Réseau sémantico-documentaire : nœuds et relations



# Réseau sémantico-documentaire : nœuds et relations



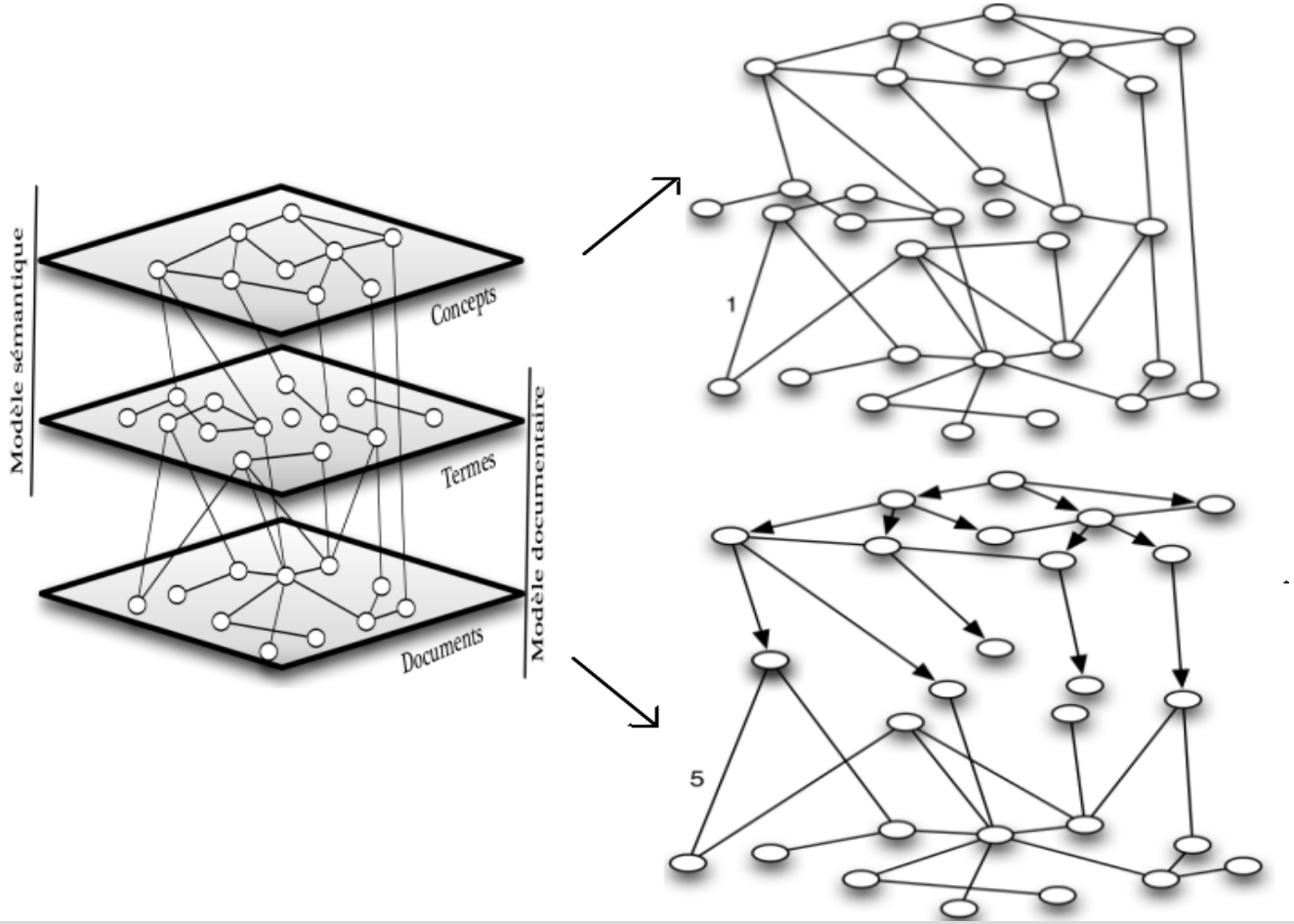
# Réseau sémantico-documentaire : exemple de document



# Traduction : graphe pondéré

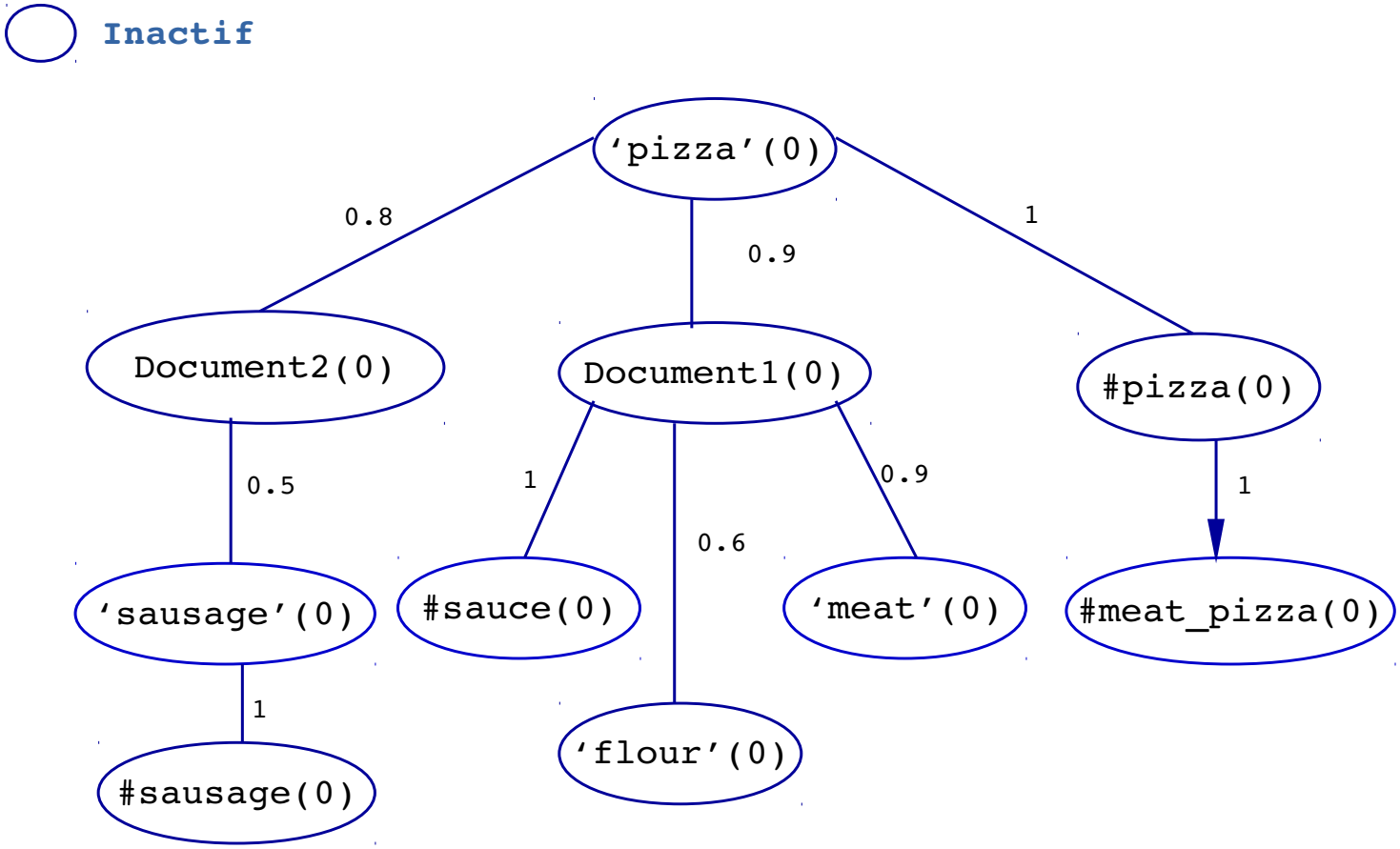
Un graphe pondéré est constitué de :

- Un ensemble de nœuds
- Un ensemble d'arcs orientés et pondérés



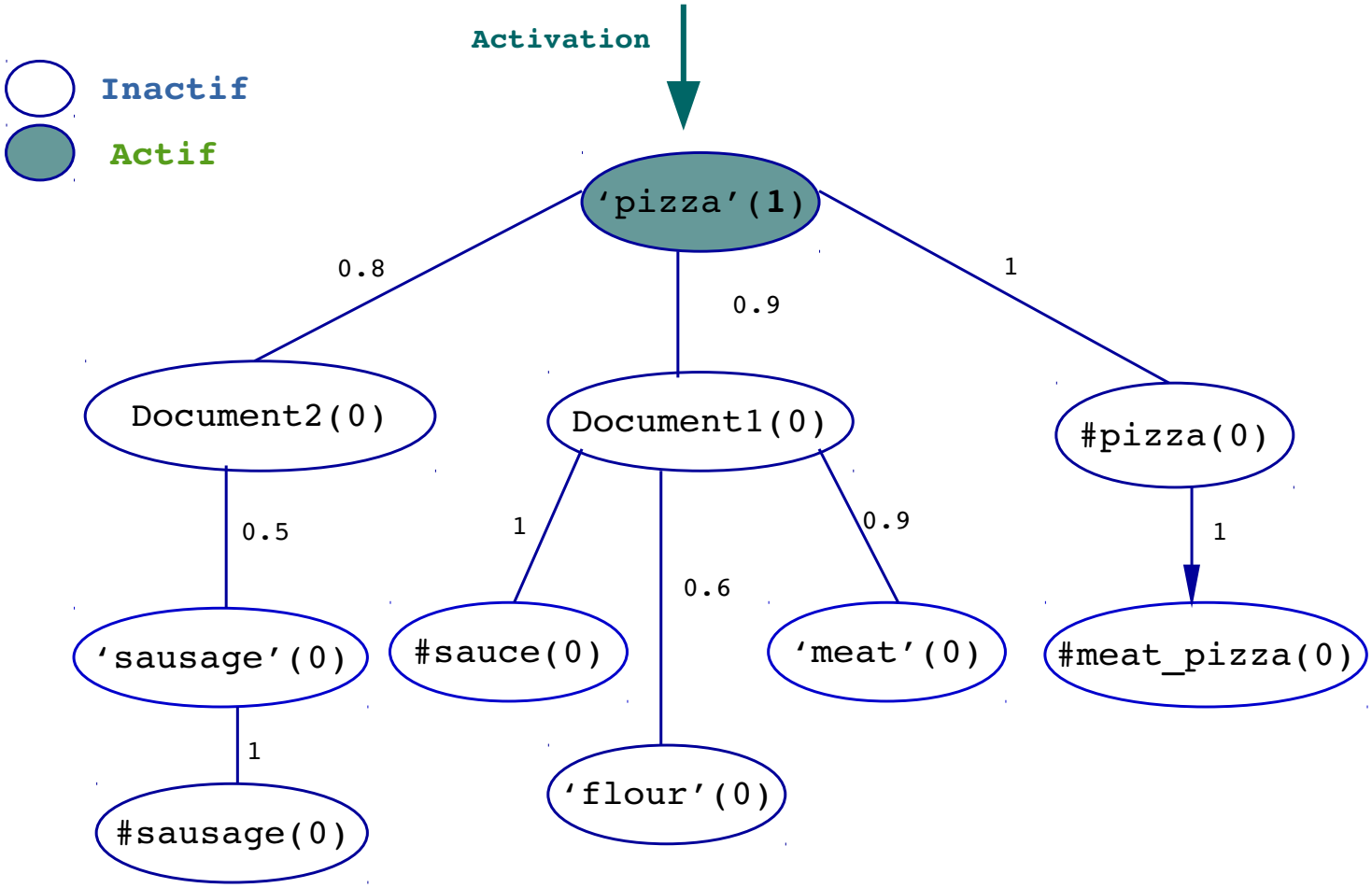
# Propagation d'activation

## Mécanisme



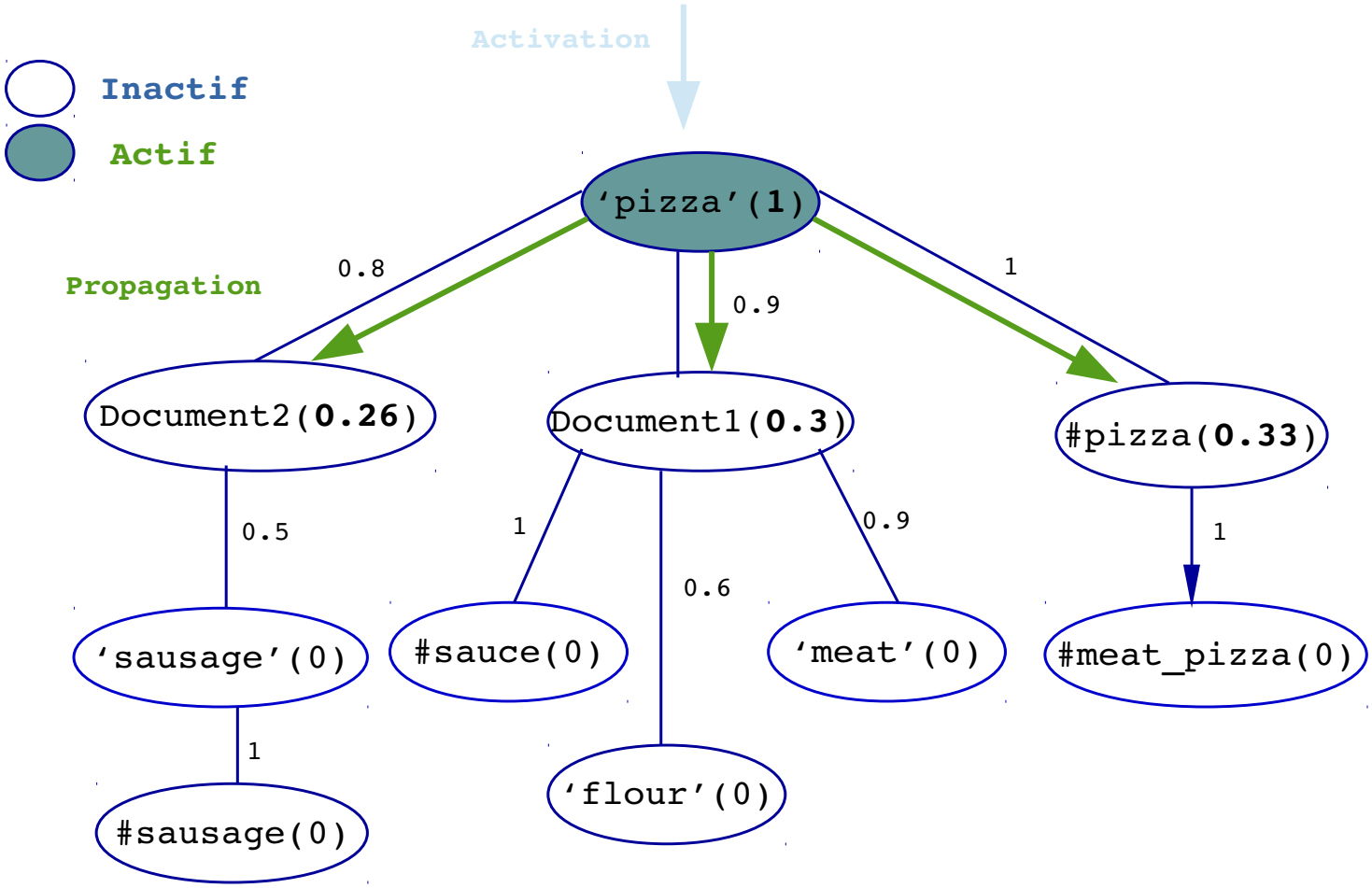
# Propagation d'activation

## Mécanisme



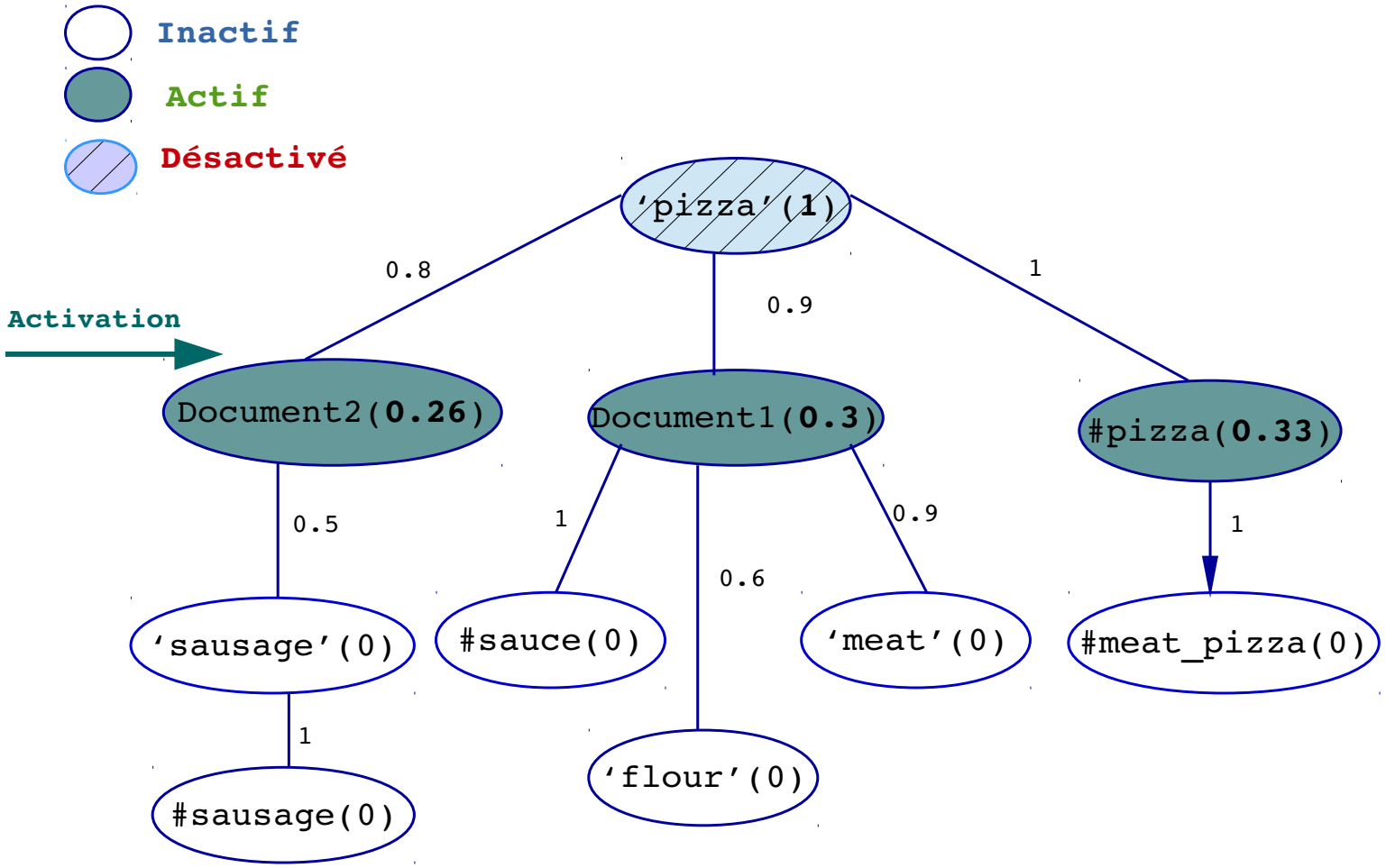
# Propagation d'activation

## Mécanisme



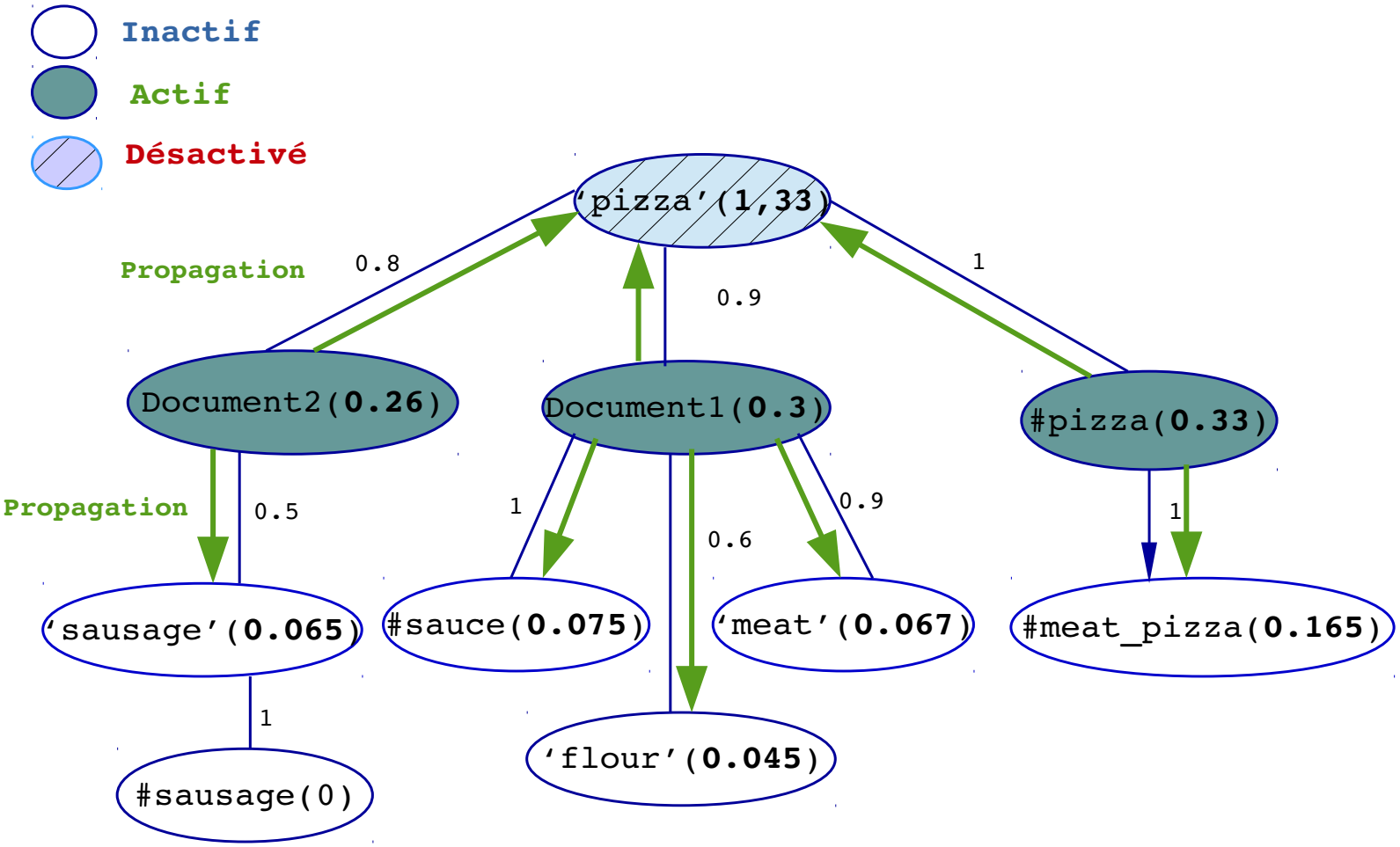
# Propagation d'activation

## Mécanisme



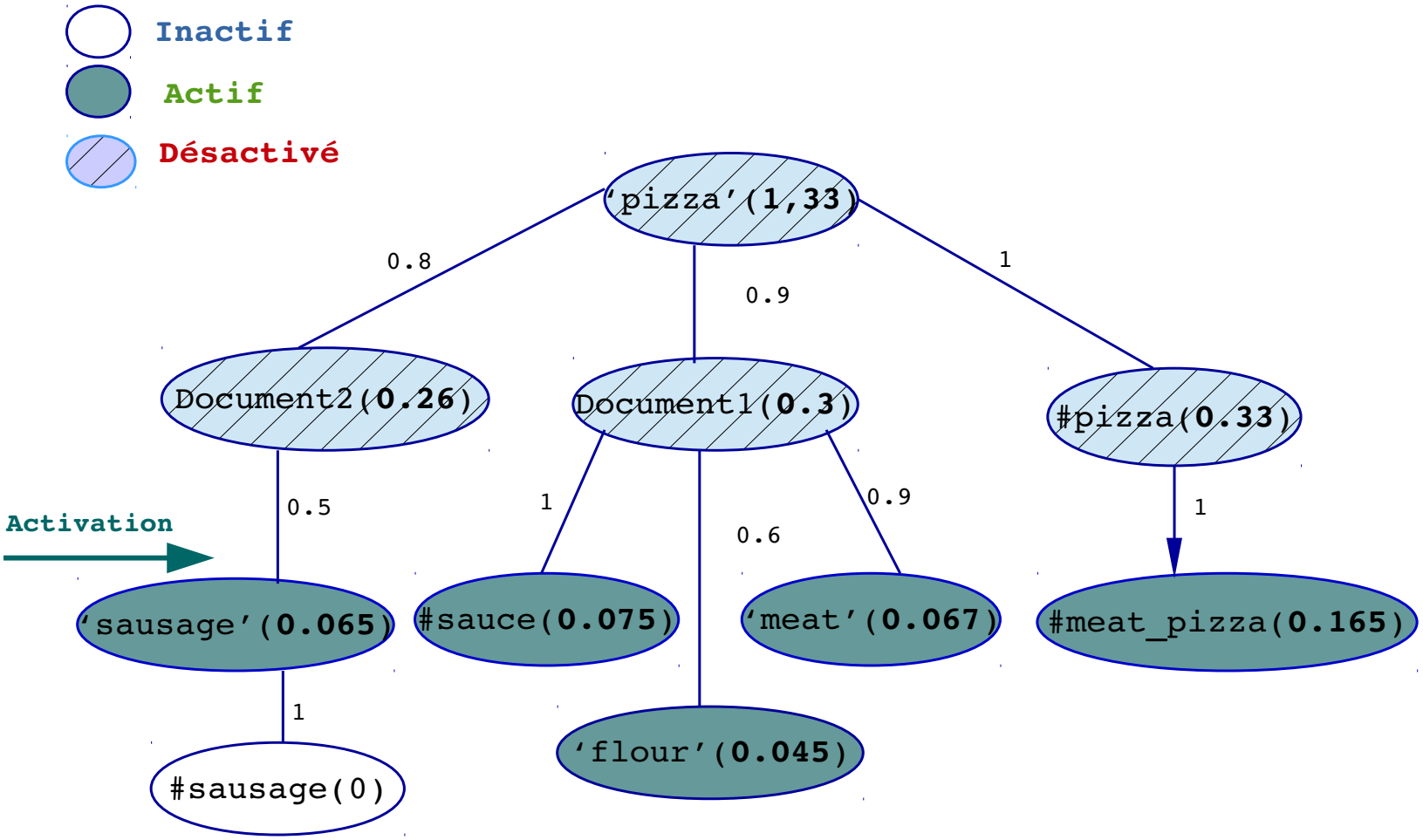
# Propagation d'activation

## Mécanisme



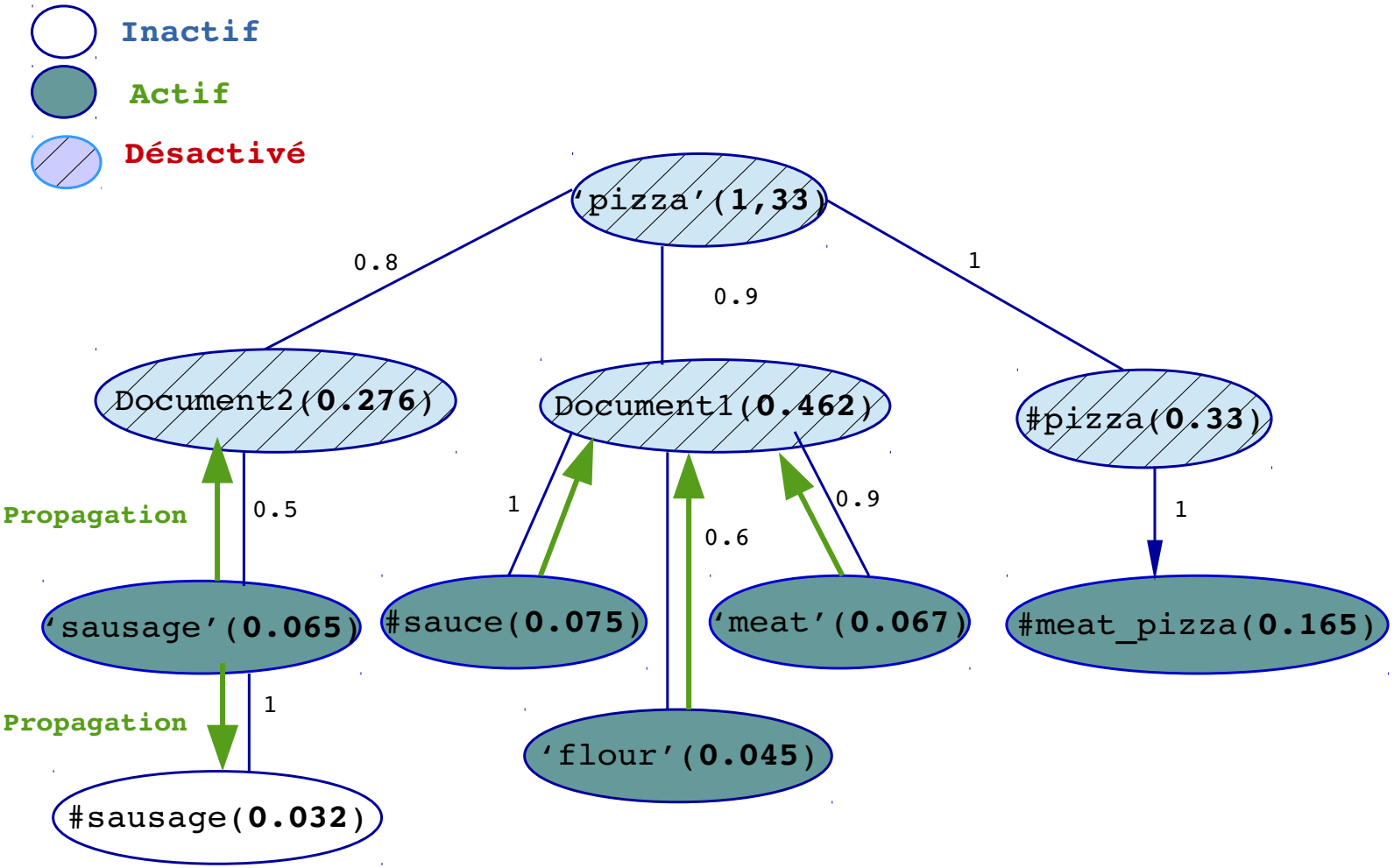
# Propagation d'activation

## Mécanisme



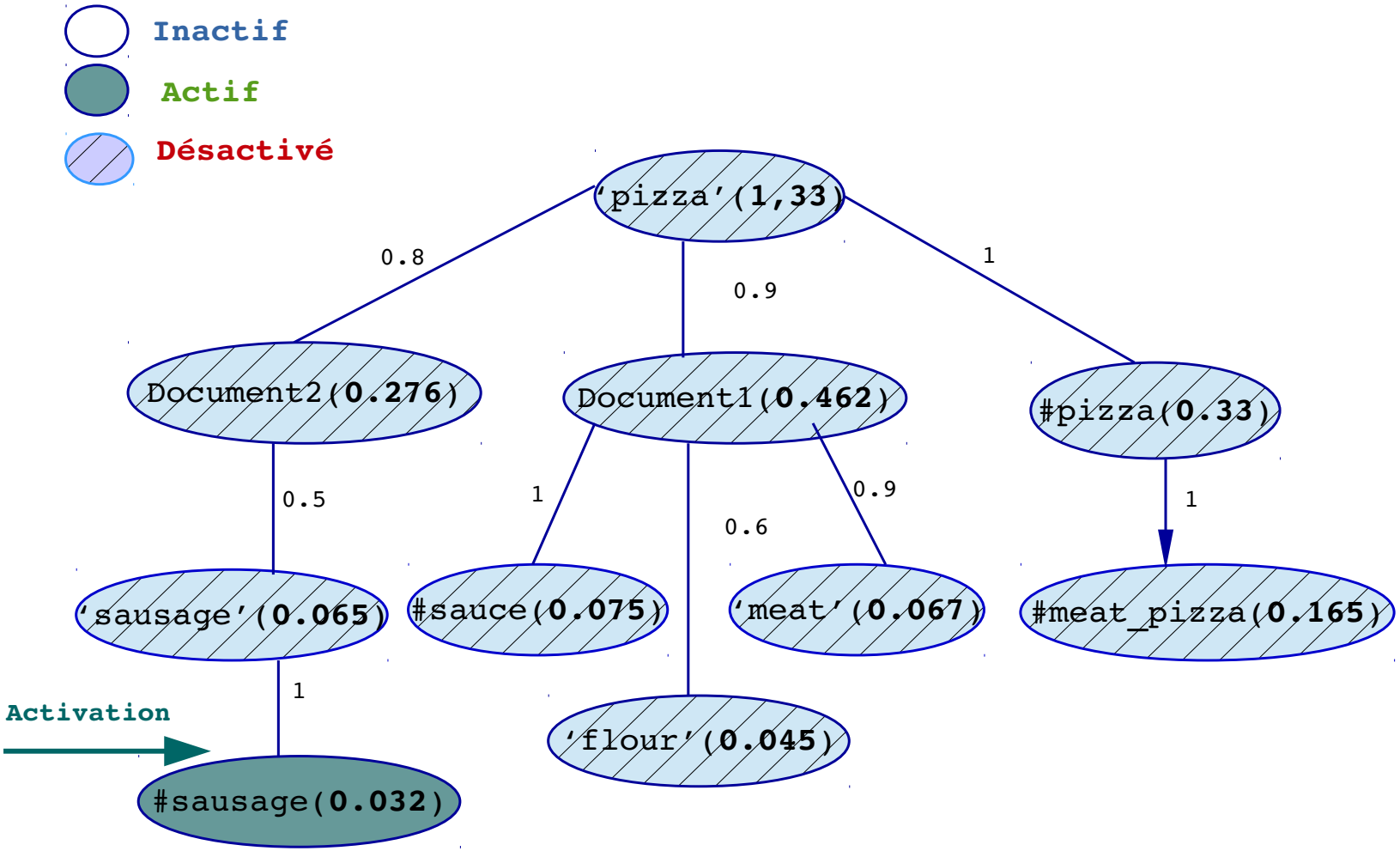
# Propagation d'activation

## Mécanisme



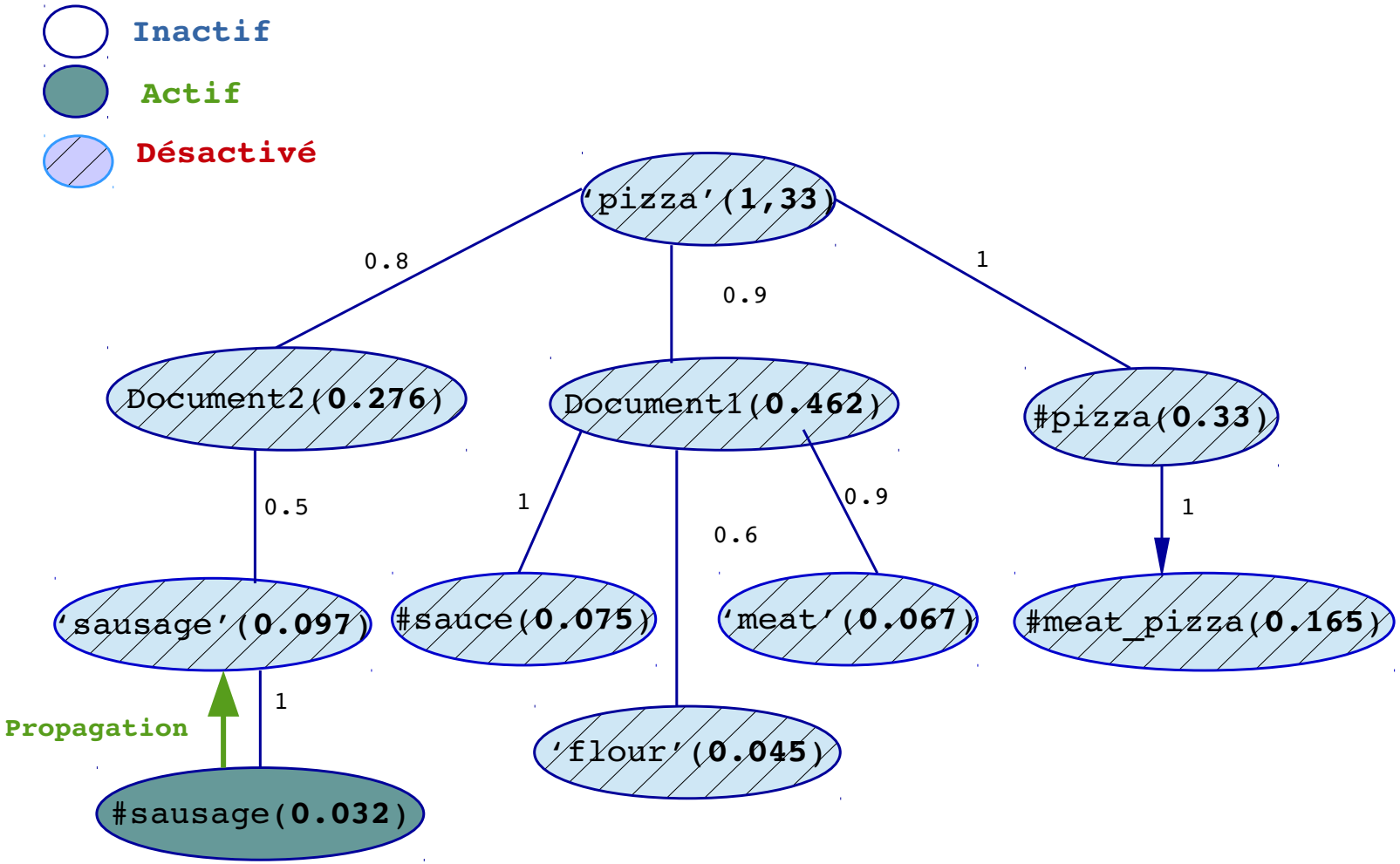
# Propagation d'activation

## Mécanisme



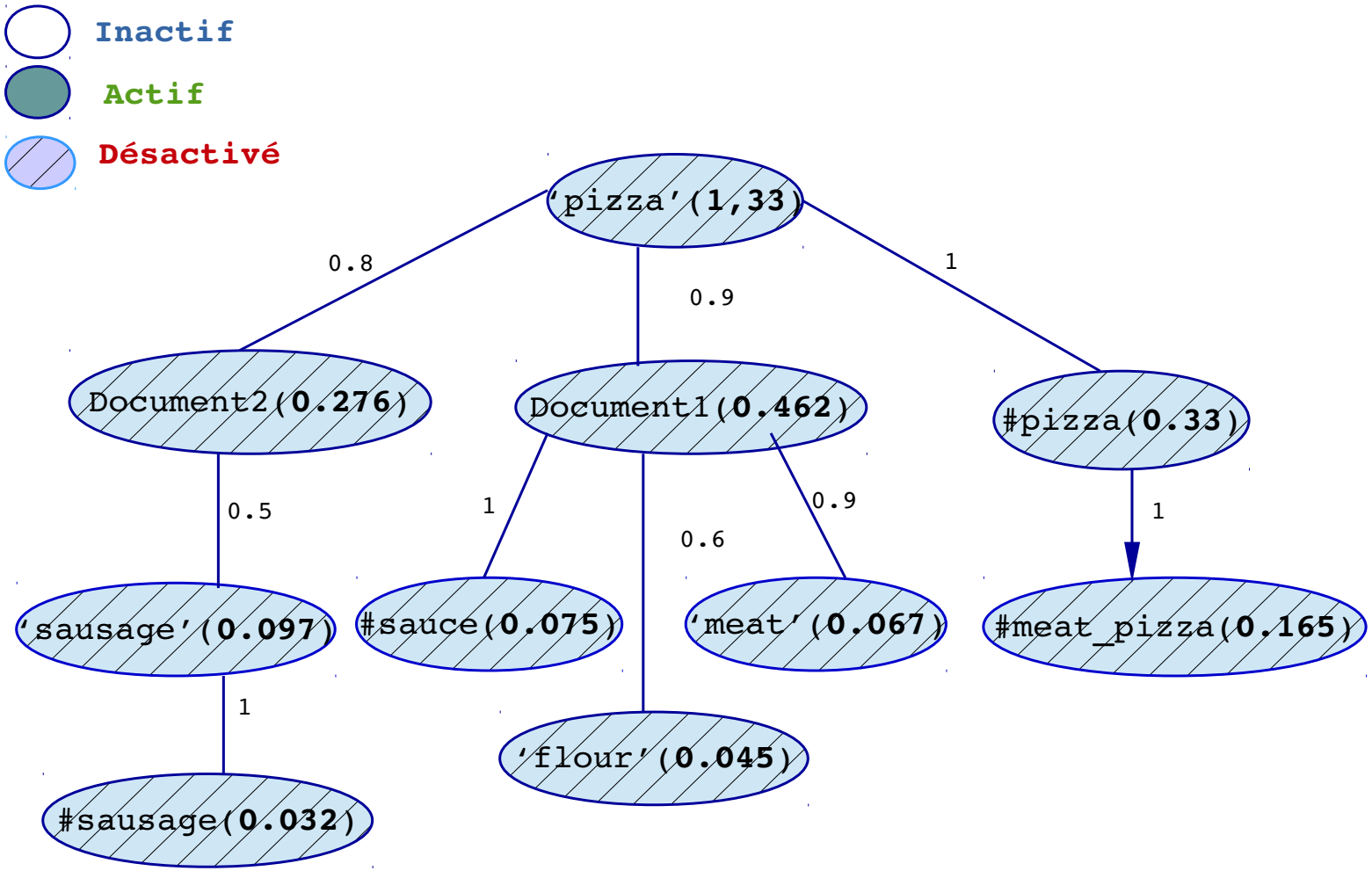
# Propagation d'activation

## Mécanisme



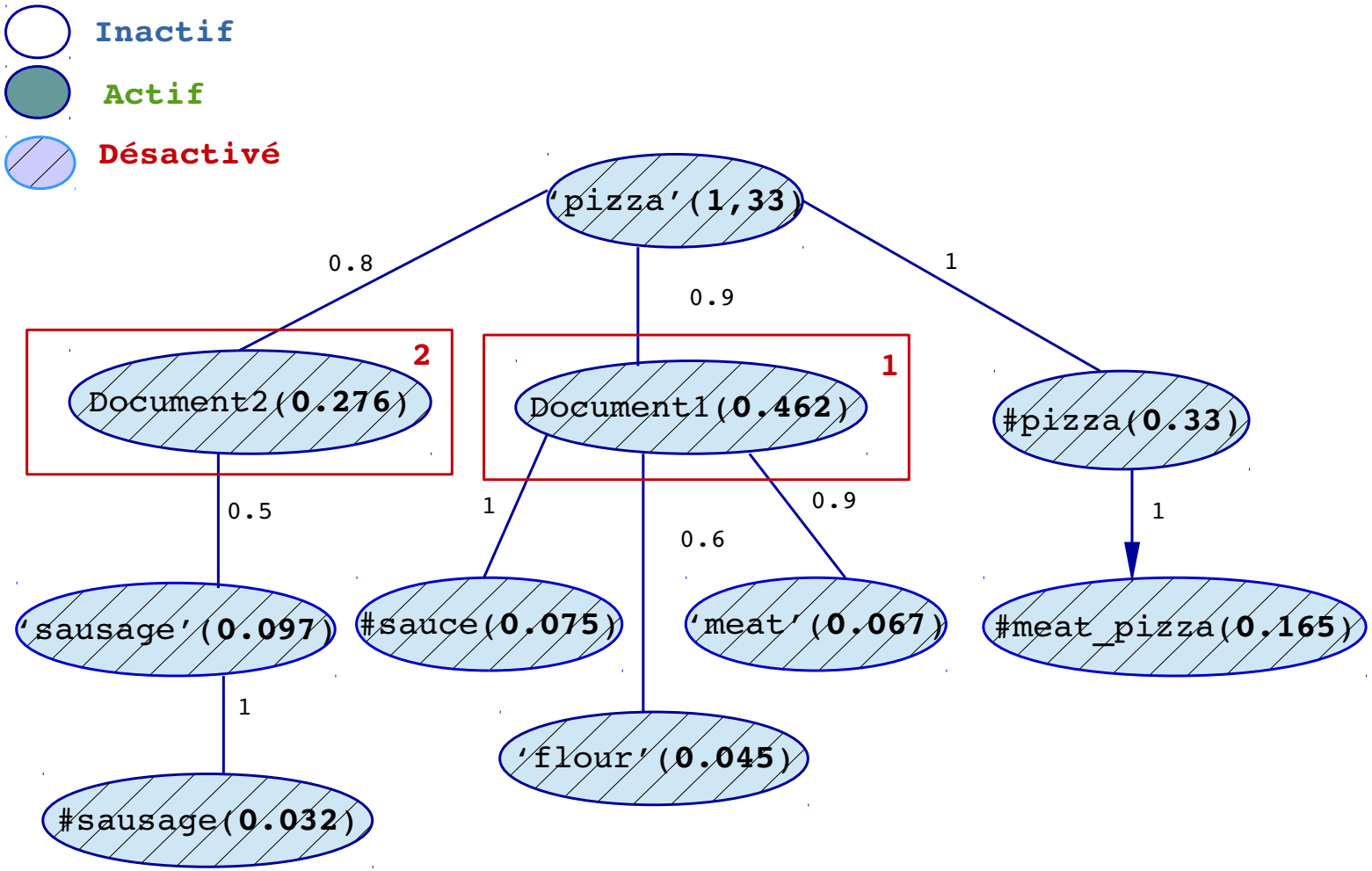
# Propagation d'activation

## Mécanisme



# Propagation d'activation

## Mécanisme



# Propagation d'activation

## Fonction d'activation : calcul de valeur d'activation

Valeur d'activation d'un nœud $i$ à une itération $k$ :

$$a_k(i) = a_{k-1}(i) + \sum_{j \in pred(i) \cap actif(k-1)} a_{k-1}(j) * w(j,i) * 1/deg(j)$$

Elle dépend de

- la valeur d'activation précédente de $i$
- la structure du graphe
- l'état des nœuds du graphe
- valeurs d'activation des voisins prédécesseurs actifs

# Propagation d'activation

## Fonction d'activation : calcul de valeur d'activation

Valeur d'activation d'un nœud $i$ à une itération $k$ :

$$a_k(i) = a_{k-1}(i) + \sum_{j \in pred(i) \cap actif(k-1)} a_{k-1}(j) * w(j,i) * 1/deg(j)$$

Elle dépend de

- la valeur d'activation précédente de $i$
- la structure du graphe
- l'état des nœuds du graphe
- valeurs d'activation des voisins prédécesseurs actifs

# Propagation d'activation

## Fonction d'activation : calcul de valeur d'activation

Valeur d'activation d'un nœud $i$ à une itération $k$ :

$$a_k(i) = a_{k-1}(i) + \sum_{j \in pred(i) \cap actif(k-1)} a_{k-1}(j) * w(j,i) * 1/deg(j)$$

Elle dépend de

- la valeur d'activation précédente de $i$
- la structure du graphe
- l'état des nœuds du graphe
- valeurs d'activation des voisins prédécesseurs actifs

# Propagation d'activation

## Fonction d'activation : calcul de valeur d'activation

Valeur d'activation d'un nœud $i$ à une itération $k$ :

$$a_k(i) = a_{k-1}(i) + \sum_{j \in pred(i) \cap actif(k-1)} a_{k-1}(j) * w(j,i) * 1/deg(j)$$

Elle dépend de

- la valeur d'activation précédente de $i$
- la structure du graphe
- l'état des nœuds du graphe
- valeurs d'activation des voisins prédécesseurs actifs

# Propagation d'activation

## Fonction d'activation : calcul de valeur d'activation

Valeur d'activation d'un nœud $i$ à une itération $k$ :

$$a_k(i) = a_{k-1}(i) + \sum_{j \in pred(i) \cap actif(k-1)} a_{k-1}(j) * w(j,i) * 1/deg(j)$$

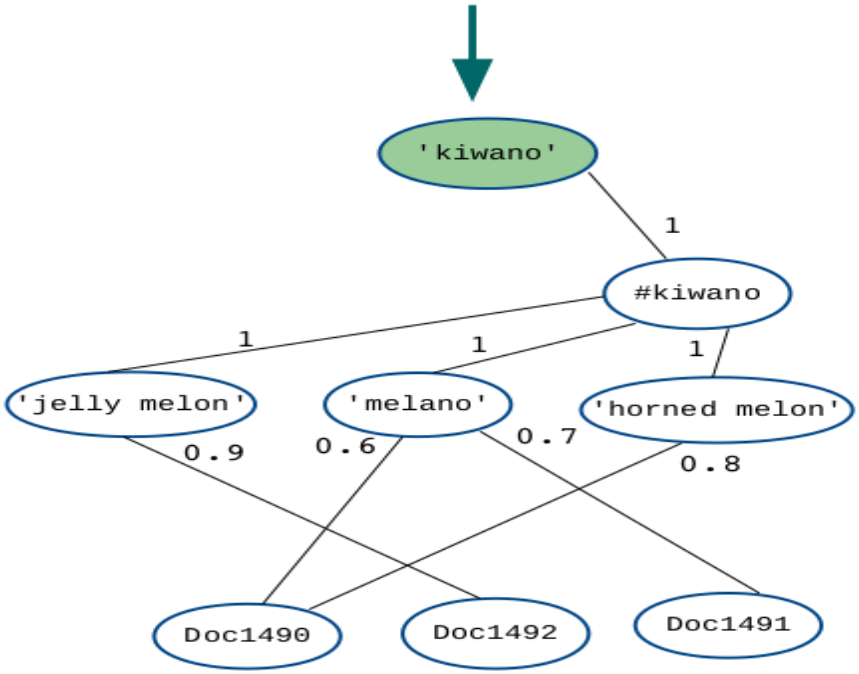
Elle dépend de

- la valeur d'activation précédente de $i$
- la structure du graphe
- l'état des nœuds du graphe
- valeurs d'activation des voisins prédécesseurs actifs

# Sémantique explicite par les classes sémantiques

## Exploitation des liens terminologiques

Graphe pondéré avec classes sémantiques : *DTC*



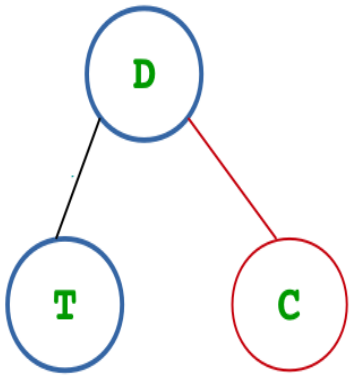

L'exploitation des liens terminologiques permet de :

- résoudre le problème de *term mismatch* : le rappel ↗
- traiter la synonymie : précision ↗, rappel ↗
- étendre la couverture sémantique sans dégrader la RIS en RI

# Sémantique explicite par les classes sémantiques

## Exploitation des liens d'annotation

Graphe pondéré avec classes sémantiques : *TDC*




L'exploitation des liens d'annotation permet de :

- poser des requêtes par des concepts de catégorisation : traiter la synonymie
- poser des requêtes par termes et concepts :
  - pallier au manque de couverture sémantique
  - traiter les problèmes de *term mismatch*

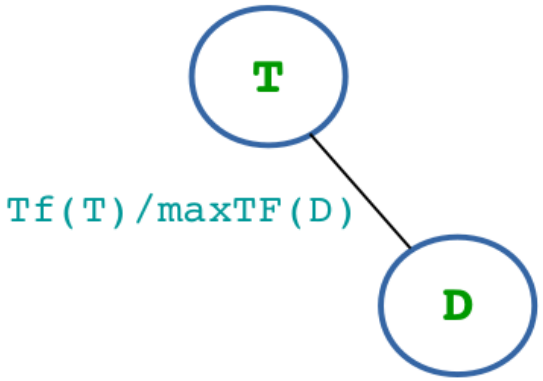
# Environnement et dispositif expérimental

- **Cadre de test**
  - Collection médicale Ohsumed87 : 54 710 documents
  - 63 requêtes Ohsumed + jugements de pertinence
  - MeSH (MEdical Subject Heading) : comporte 40 247 concepts
  - Annotation manuelle au regard d'une partie de MeSH
- **Plate-formes et logiciels**
  - Terrier SIR [Bannour and Zargayouna, 2012] + JUNG
  - Annotateur MetaMap [Aronson, 2001] + désambiguïsation par défaut
- **Métriques d'évaluation** : courbes de Rappel-Précision, R-PREC
- **Graphes pondérés :**
  - Graphe sans sémantique *TD*
  - Graphe avec classes sémantiques et liens d'annotation *TDC*
- **Nombre d'itérations :** entre 6 et 8 itérations

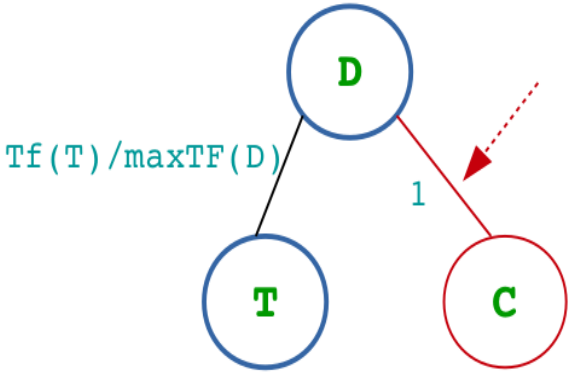
# Impact de la sémantique

## Modèle **TD** vs. modèle **TDC**

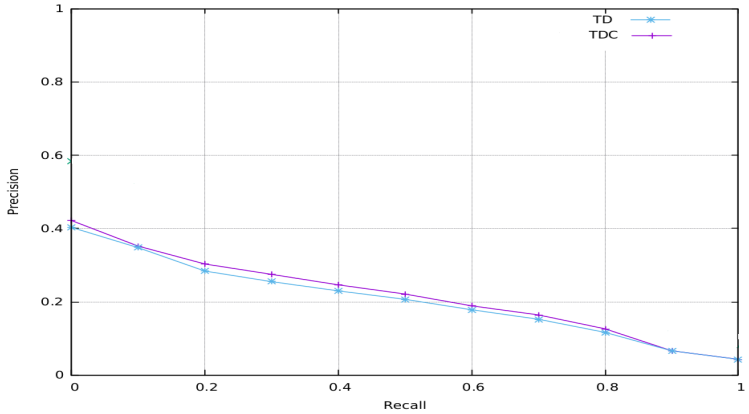
- Interrogation par termes
- Modèle Terme/Document avec fréquence normalisée




- Modèle Terme/Document/Concept avec classes sémantiques



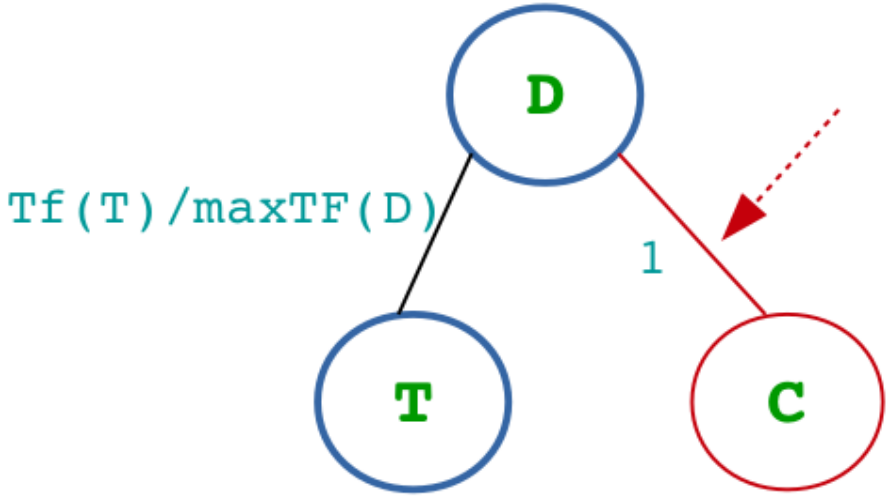
# Imapct de la sémantique




- Résultats quasi-stable entre le modèle **TD** et le modèle **TDC**, avec une légère amélioration de 1% de la R-PREC
  - ✓ Amélioration des classements des documents pertinents pour quelques requêtes : l'impact de la co-occurrence notamment des concepts
  - ✗ Relâchement rapide des valeurs d'activation avec la distance : documents découverts par la co-occurrence et la synonymie mal classés
  - ✗ Pas de résolution du *term mismatch*
    $\rightarrow$ **solution :** interrogation double par les termes et les concepts + ontologie couvrante

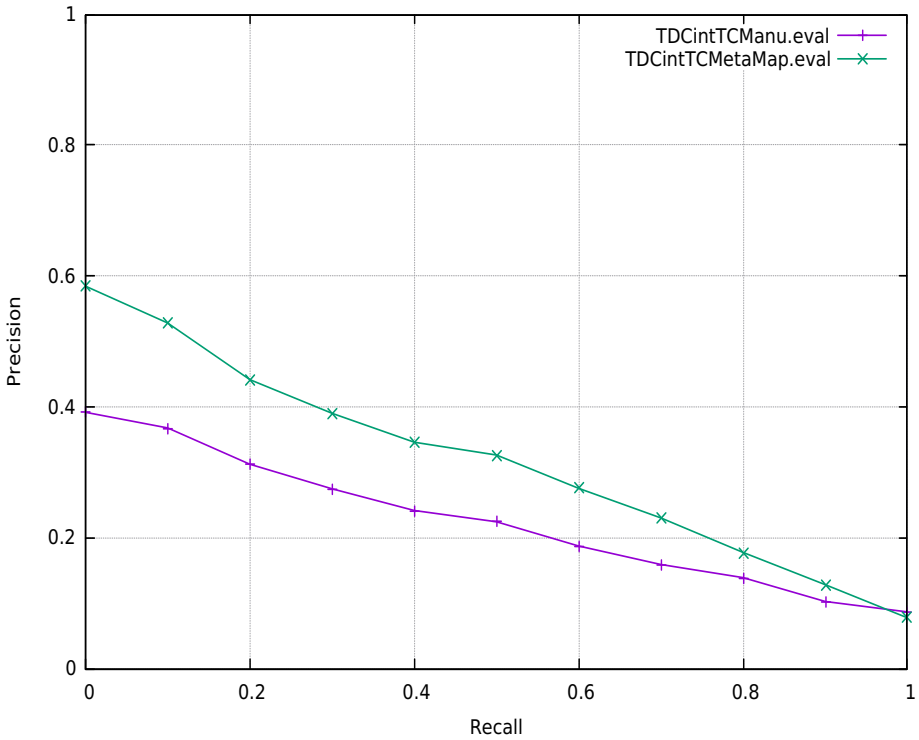
# Impact de la densité du graphe

## TDCintTC$_{MetaMap}$ vs. TDCintTC$_{Manu}$




- Interrogation par les termes et les concepts **TC**
- Annotation par MetaMap : ontologie MeSH couvrante, proche du texte
- Annotation manuelle : partie de l'ontologie MeSH, catégorisation

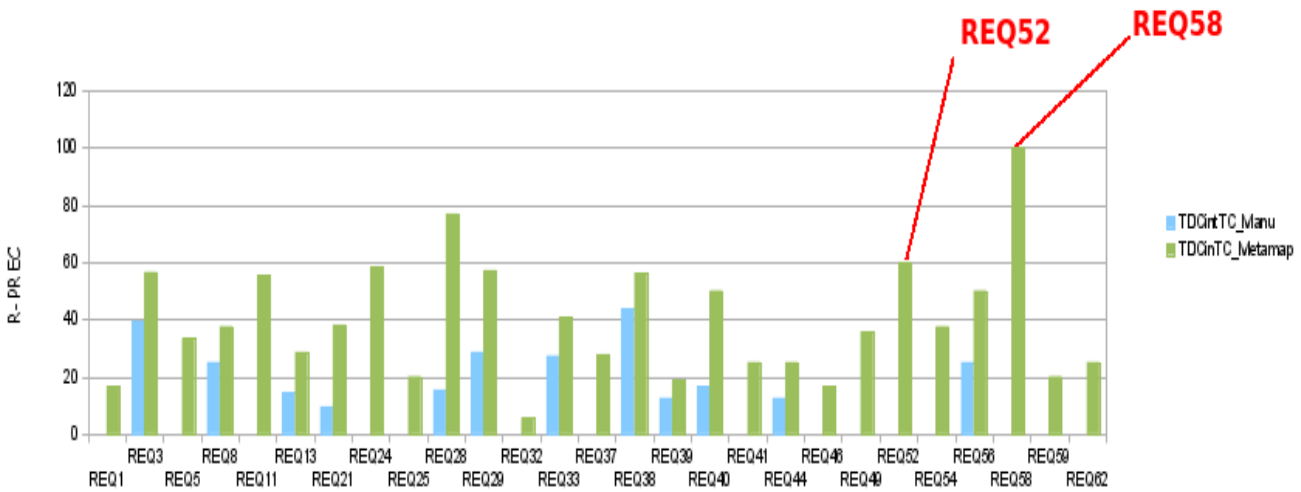
# Impact de la densité du graphe




- L'annotation par MetaMap (**TDCintTC**$_{MetaMap}$) donne de meilleurs résultats par rapport à l'annotation manuelle (**TDCintTC**$_{Manu}$)
  - Requêtes avec amélioration : 27
  - Requêtes avec stabilisation : 25
  - Requêtes avec dégradation : 10

# Impact de la densité du graphe

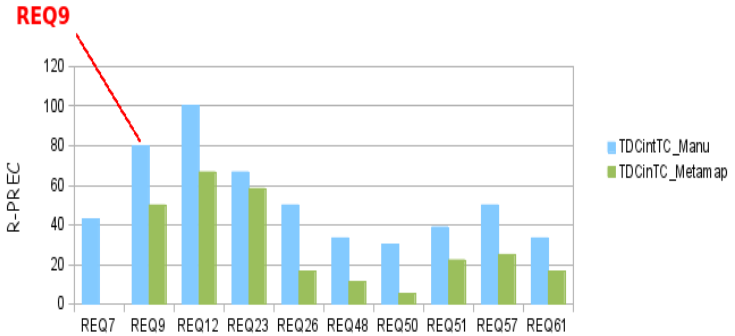
## Analyse des résultats améliorés




- Couverture de l'ontologie exploitée par MetaMap
  - **Exemple :** *Requête 58* : « femme avec un dos thoracique, maladie **Scheurmann** »
  - l'annotation manuelle n'a pas permis d'annoter avec le concept `#scheuermann's_disease` que donne MetaMap
- Précision de l'annotation automatique
  - **Exemple :** *Requête 52 : « une personne de 60 ans avec **abcès pulmonaire**... »*
  - annotation manuelle : `#lung`, `#lung_disease`, `#lung_absces`
  - annotation automatique : `#Lung_Absces`

# Impact de la densité du graphe

## Analyse des résultats dégradés




- Erreurs d'annotation de MetaMap
- Non prise en compte des concepts proches
- **Exemple :** *Requête 9 : « femme **enceinte de 3 mois**, une **rh isoimmunisation** »*
  - MetaMap annote la requête
    - le terme "rh" : `#Rh Isoimmunization`
    - le terme "enceinte" : `#Gravidity`
  - MetaMap annote les documents pertinents
    - le terme "rh" : `#Macaca mulatta`, `#Rhodium`
    - le terme "enceinte" : `#Pregnancy`

→ **solution :** exploitation de la structure de l'ontologie

# Conclusion et perspectives

## Conclusion

- ✓ Modèle dédié à la recherche d'information sémantique :
  - réseau sémantico-documentaire + traductions en graphe pondéré
  - appariement par propagation d'activation + interrogation par différents points d'entrée
- ✓ Expérimentations dans le domaine médical :
  - Apport des classes sémantiques et des liens d'annotation
  - Impact de la couverture sémantique
  - Impact du type d'annotation

# Conclusion et perspectives

## Conclusion

✓ Modèle dédié à la recherche d'information sémantique :
- réseau sémantico-documentaire + traductions en graphe pondéré
- appariement par propagation d'activation + interrogation par différents points d'entrée

✓ Expérimentations dans le domaine médical :
- Apport des classes sémantiques et des liens d'annotation
- Impact de la couverture sémantique
- Impact du type d'annotation

## Perspectives

- Résoudre les limites du modèle :
  - ✗ Relâchement des valeurs d'activation avec la distance
  - ✗ Non exploitation de la structure de l'ontologie : erreurs d'annotation

  → Réajustement des mécanismes de contrôle
- Conduire plus d'expérimentations : autres traductions en graphe pondéré , impact d'autres paramétrages (ex. poids des arcs), liens hiérarchiques et du domaine

# MERCI POUR VOTRE ATTENTION

# Ressources sémantiques pour la RIS

- enrichissement de la représentation des documents
  - annotation sémantique
    - désambiguïsation sémantique : [Stetina et al., 1998, Baziz, 2005]
  - choix d'unités d'index
    - concepts [Gonzalo et al., 1998, Stetina et al., 1998, Khan, 2000]
    - termes et concepts [Dinh and Tamine, 2010, Hamadan et al., 2012]
- enrichissement des scores de correspondance et schémas de pondération : enrichissement des schémas classiques par des similarités sémantiques
  - pondération de concepts [Baziz, 2005, Koopman et al., 2012, Boubekeur and Azzoug, 2013]
  - pondération de termes [Zargayouna, 2005]
  - scores des documents [Gerani et al., 2012]
- expansion des requêtes [Salton, 1968, Bhogal et al., 2007]

# Propagation d'activation

## Propagation d'activation et RI

### Recherche d'information

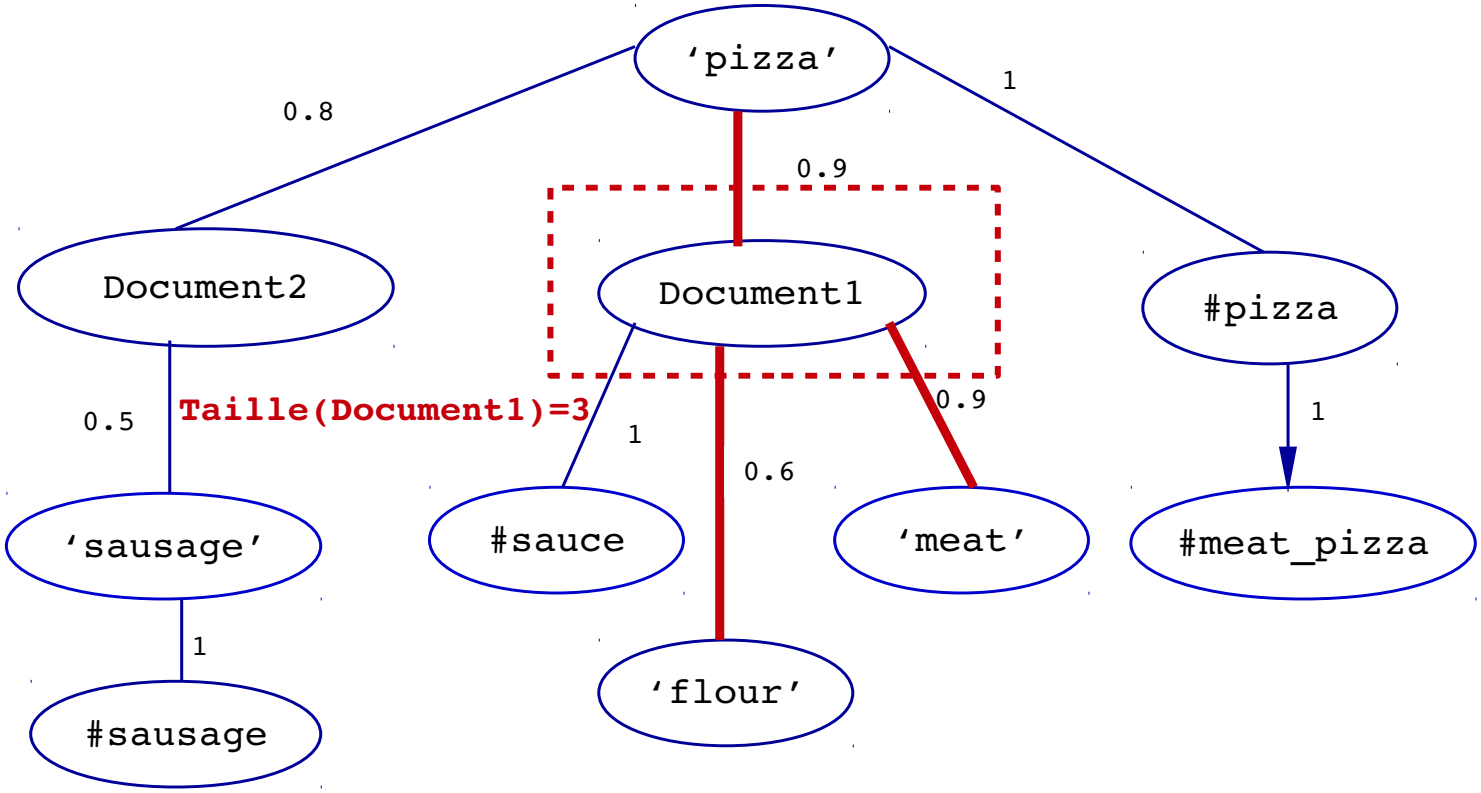
- **Objectif** : exploiter les associations entre termes et documents en RI [Crestani, 1997]
  - RI classique [Preece, 1981, Cohen and Kjeldsen, 1987, Salton and Buckley, 1988]
  - Expansion des requêtes [Shoval, 1981]
  - Recherche par l'exemple [Croft et al., 1989] (GRANT)
  - Classification [Savoy, 1992]
- **Limites**
  - Construction manuelle du réseau sémantique
  - Réglage difficile des contraintes
  - Indisponibilité des ressources sémantiques

### Web sémantique

- Travaux hybrides de la RI et du WS : [Rocha et al., 2004, Jiang and Tan, 2006, Schumacher et al., 2008]
  - Ontologie lexicalisée
  - Recherche de données non de documents

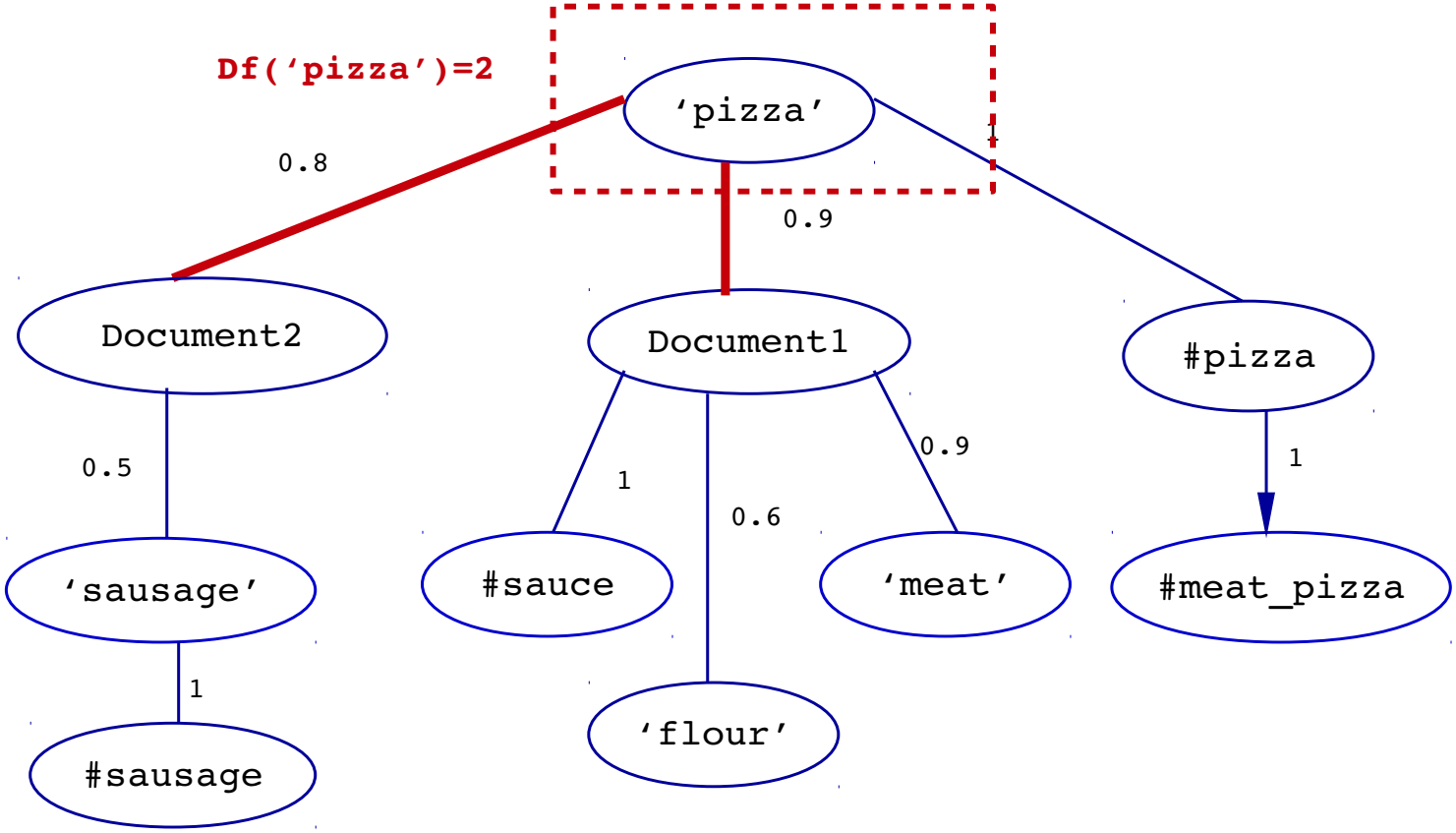
# Propagation d'activation

## Propriétés implicites



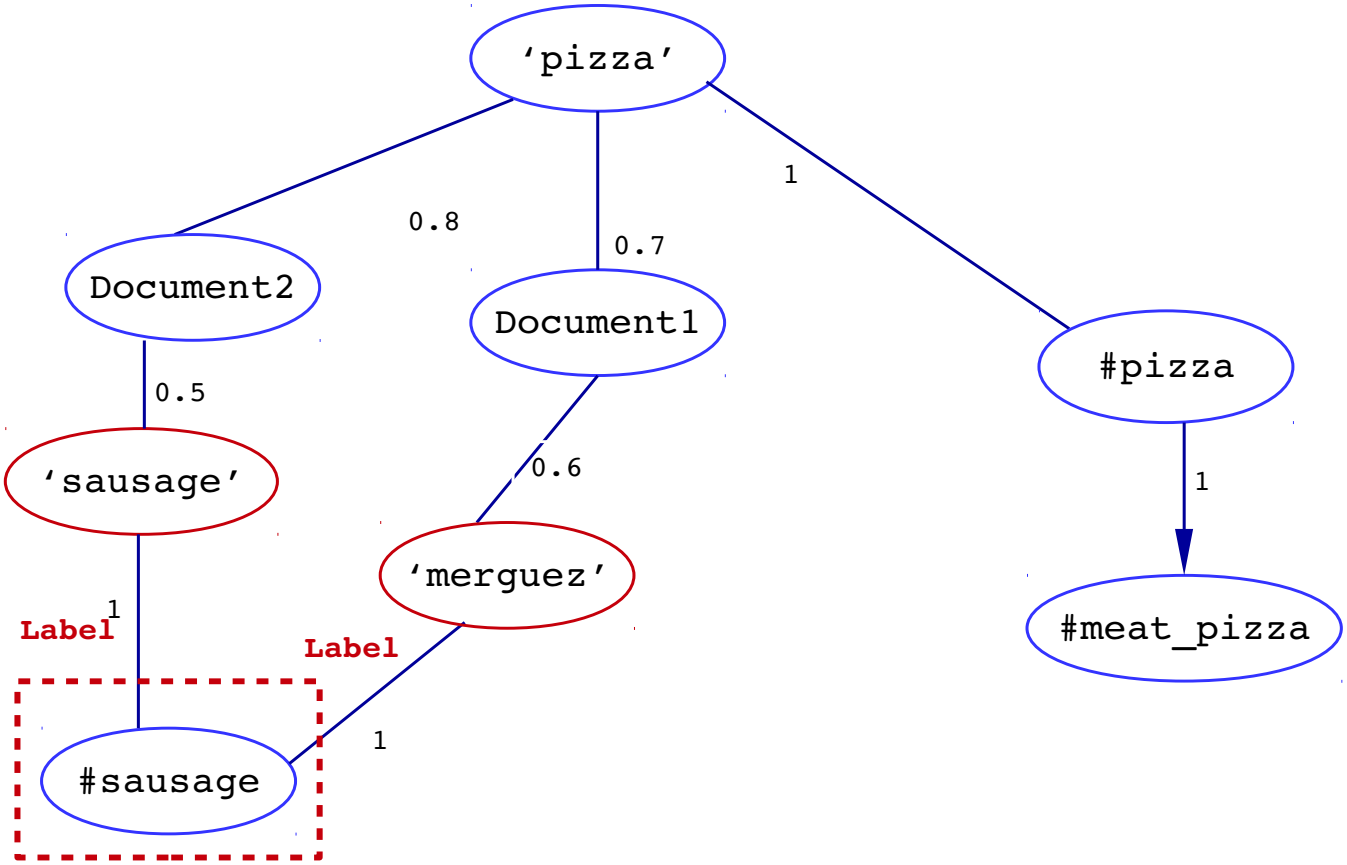
# Propagation d'activation

## Propriétés implicites



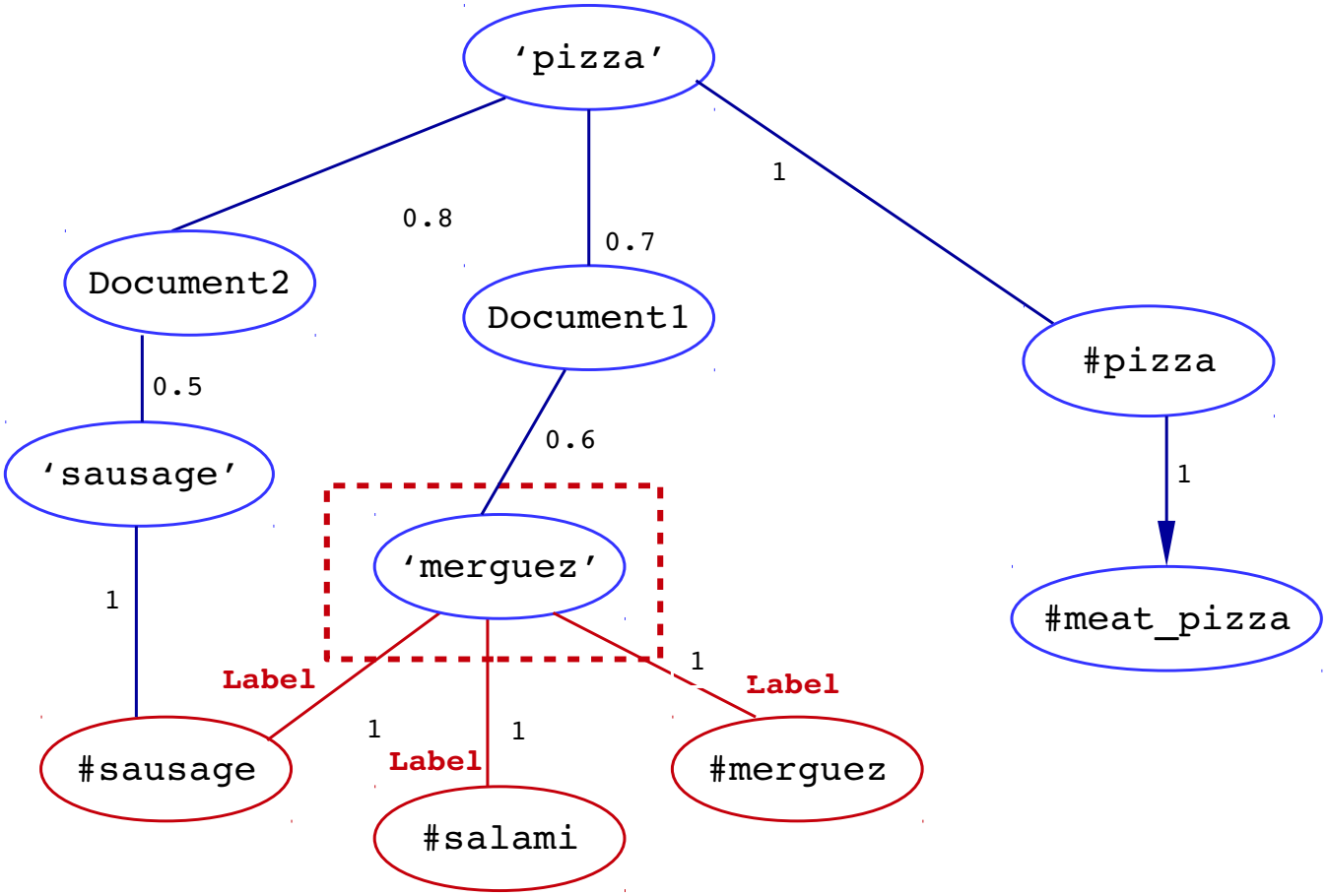
# Propagation d'activation

## Propriétés implicites



# Propagation d'activation

## Propriétés implicites



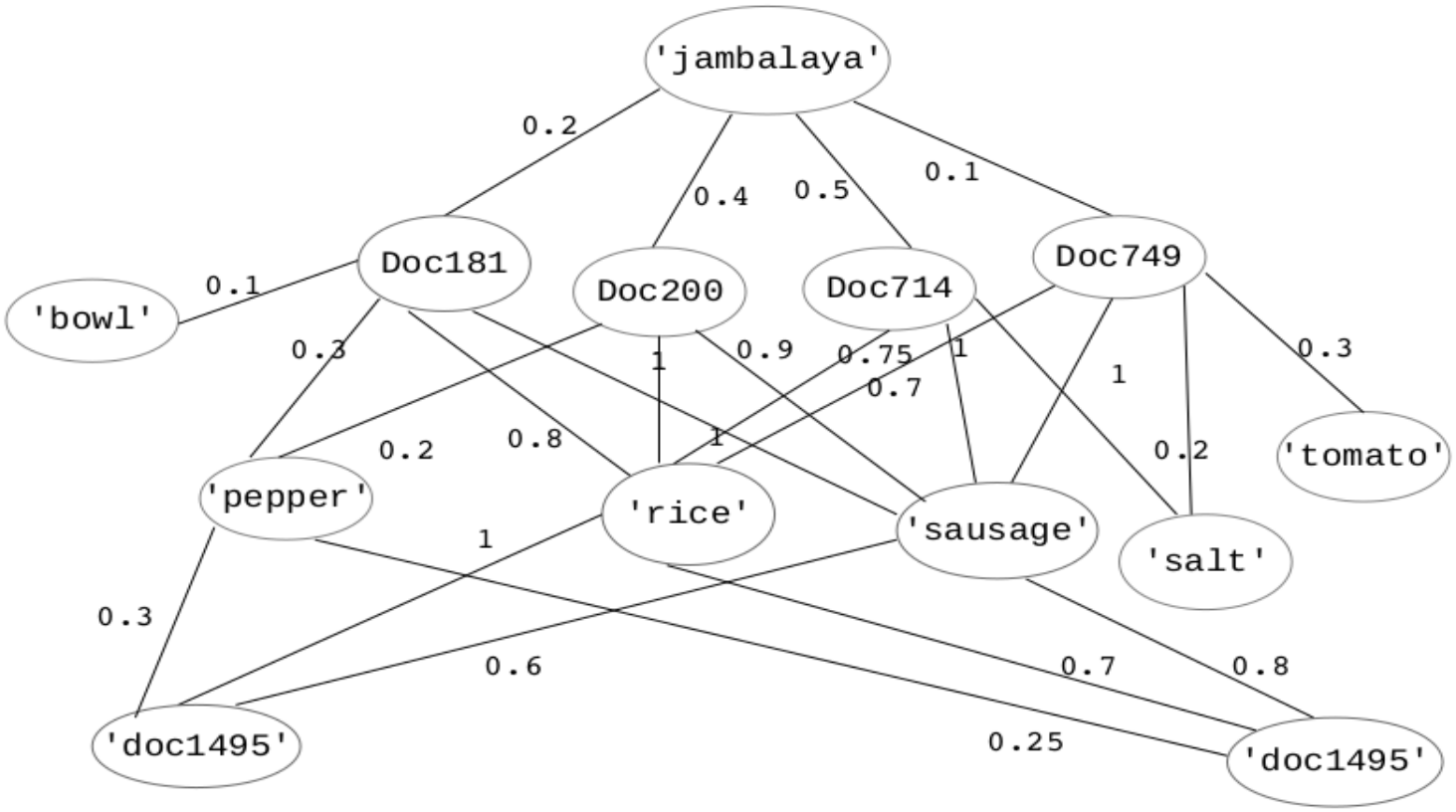
# Sémantique implicite

## Exploitation de la co-occurrence

**Définition** : deux termes sont co-occurrents s'ils apparaissent ensemble dans les mêmes documents (plus que par hasard)

- Désambiguïser le vocabulaire : la précision ↗
- Activer de nouveaux documents pertinents : le rappel ↗

REQ : *"Prepare **jambalaya**"*



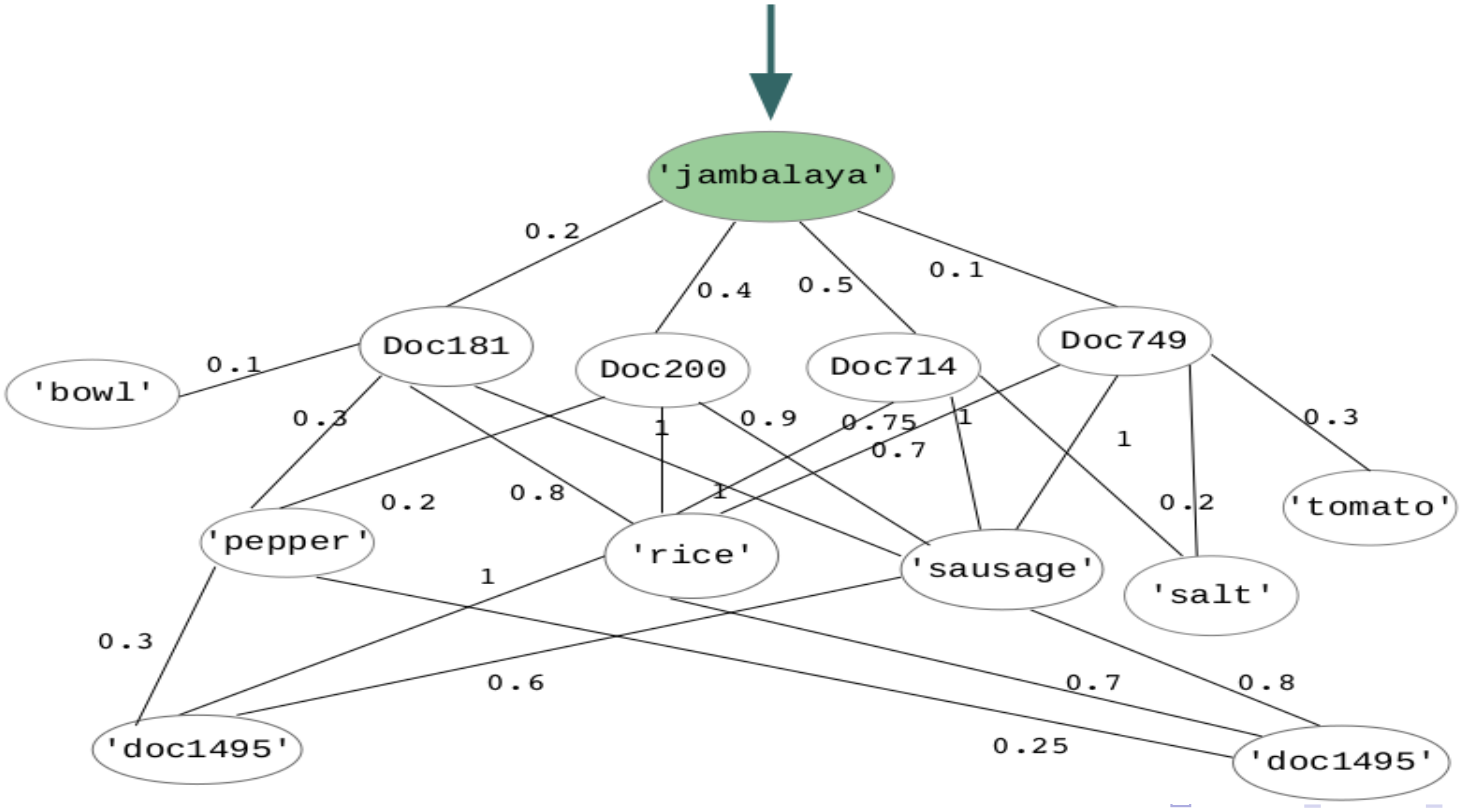
# Sémantique implicite

## Exploitation de la co-occurrence

**Définition** : deux termes sont co-occurrents s'ils apparaissent ensemble dans les mêmes documents (plus que par hasard)

- Désambiguïser le vocabulaire : la précision ↗
- Activer de nouveaux documents pertinents : le rappel ↗

REQ : *"Prepare **jambalaya**"*



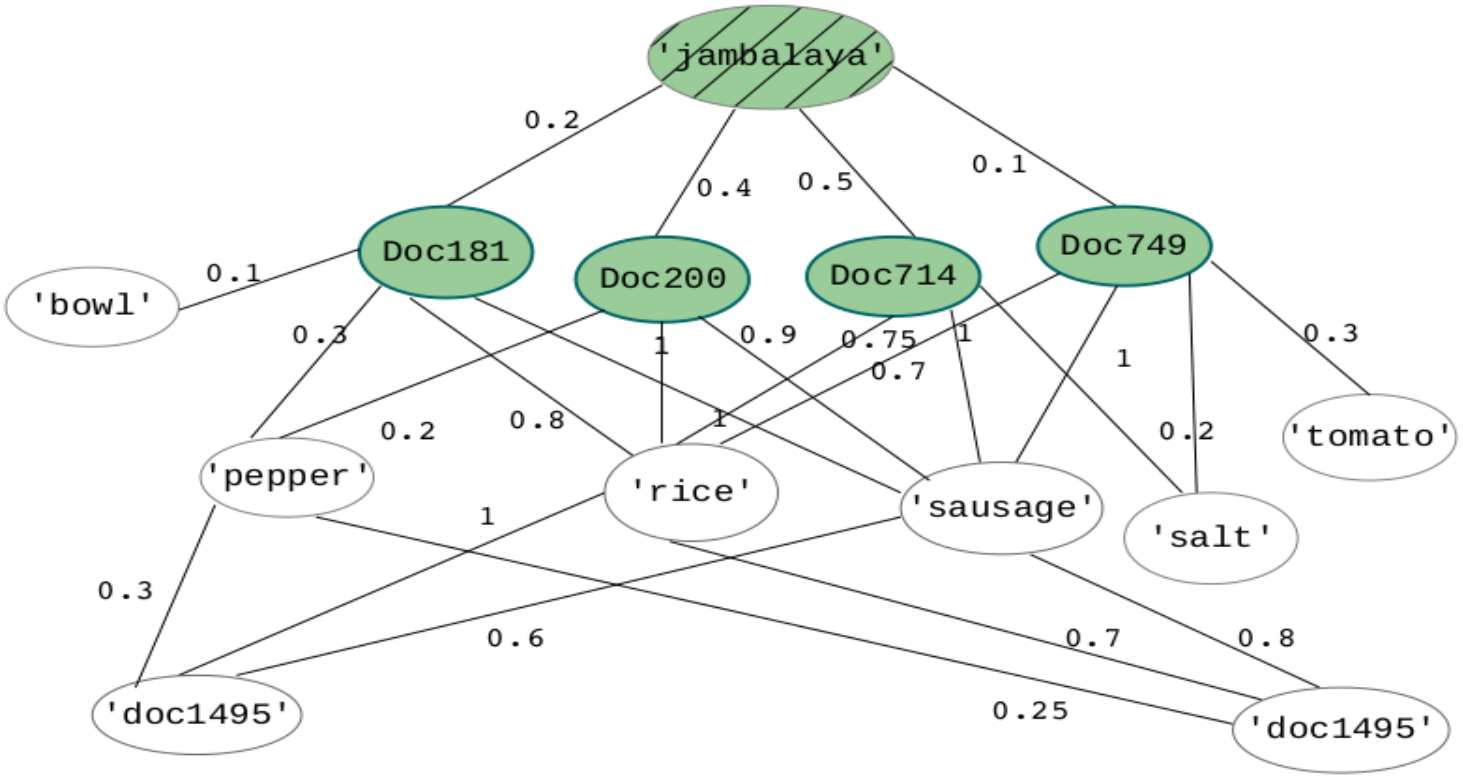
# Sémantique implicite

## Exploitation de la co-occurrence

**Définition** : deux termes sont co-occurrents s'ils apparaissent ensemble dans les mêmes documents (plus que par hasard)

- Désambiguïser le vocabulaire : la précision ↗
- Activer de nouveaux documents pertinents : le rappel ↗

REQ : *"Prepare **jambalaya**"*



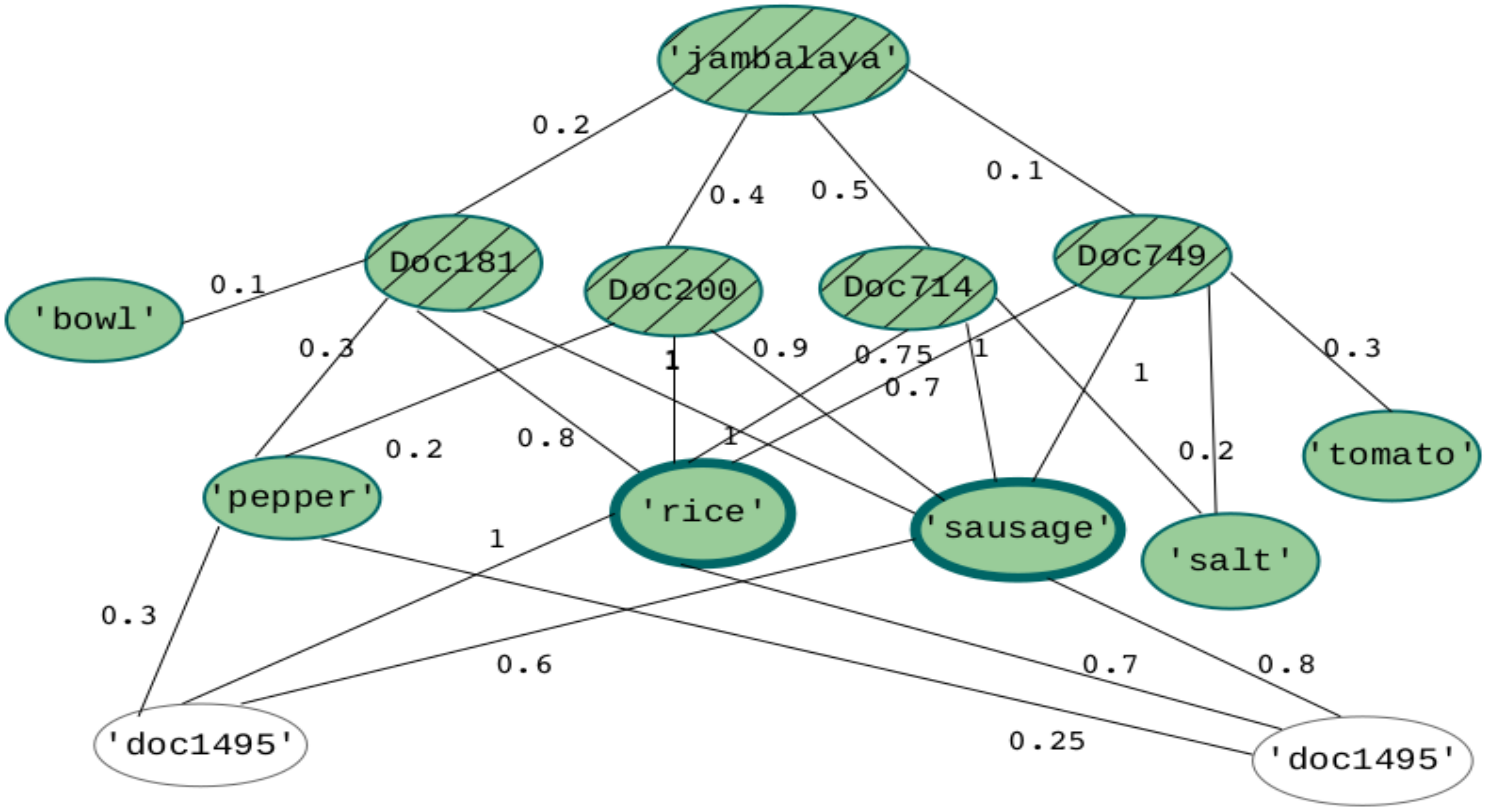
# Sémantique implicite

## Exploitation de la co-occurrence

**Définition** : deux termes sont co-occurrents s'ils apparaissent ensemble dans les mêmes documents (plus que par hasard)

- Désambiguïser le vocabulaire : la précision ↗
- Activer de nouveaux documents pertinents : le rappel ↗

REQ : *"Prepare **jambalaya**"*



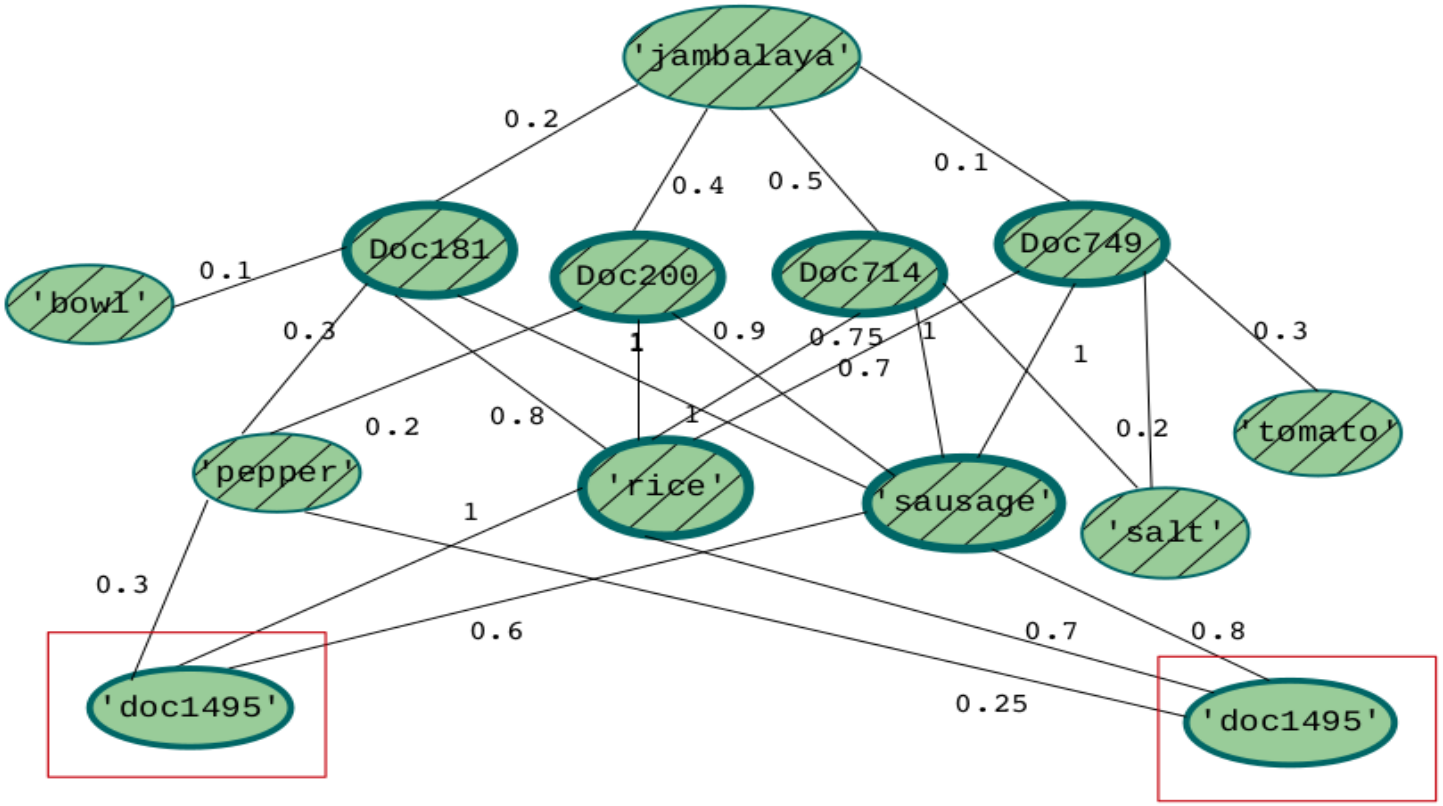
# Sémantique implicite

## Exploitation de la co-occurrence

**Définition** : deux termes sont co-occurrents s'ils apparaissent ensemble dans les mêmes documents (plus que par hasard)

- Désambiguïser le vocabulaire : la précision ↗
- Activer de nouveaux documents pertinents : le rappel ↗

REQ : *"Prepare **jambalaya**"*



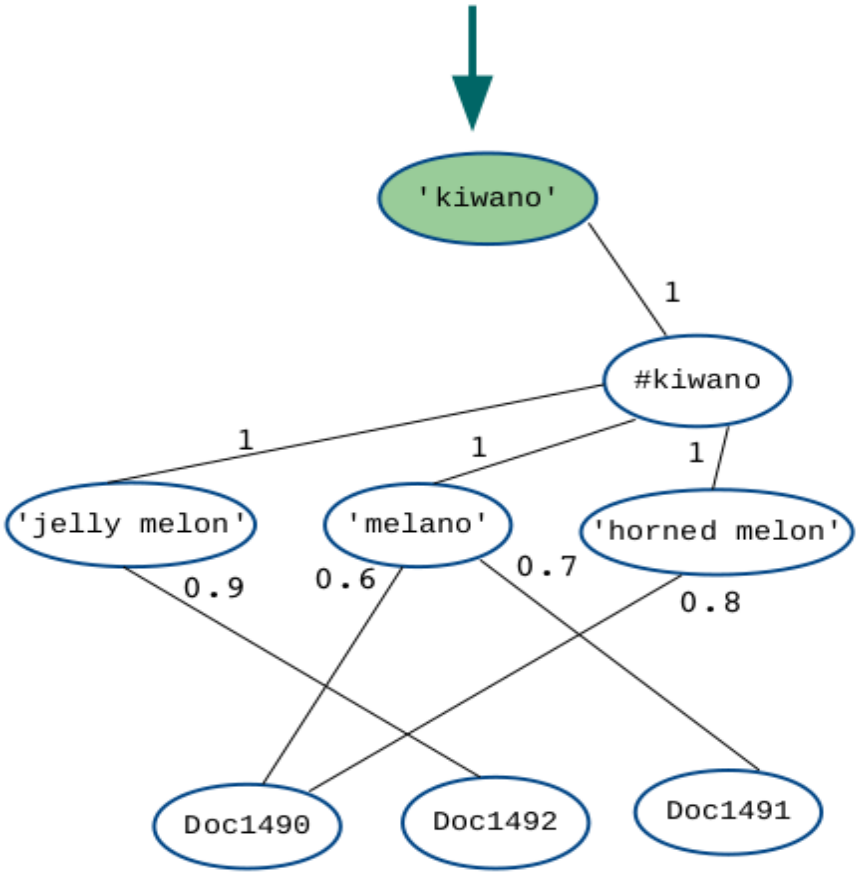
# Sémantique explicite par les classes sémantiques

## Résolution du *term mismatch*

**Définition** : un terme de la requête est absent dans les documents mais présent dans l'ontologie

- Retrouver de nouveaux documents pertinents : le rappel ↗

REQ : *"I would like a salad with **kiwano**"*



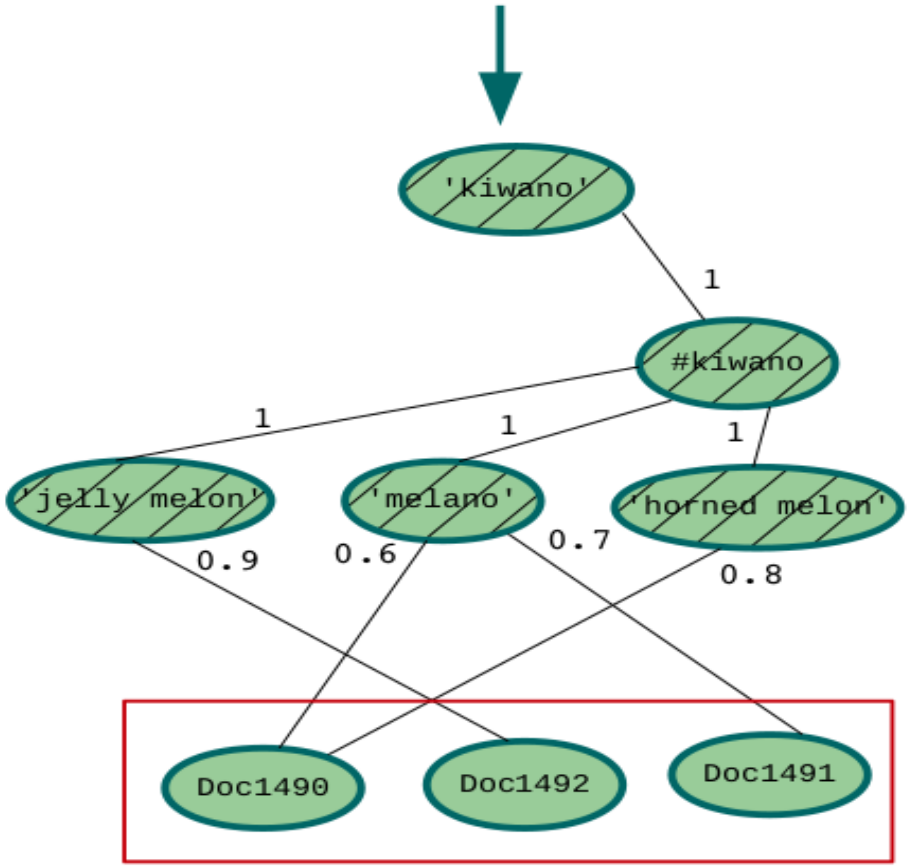
# Sémantique explicite par les classes sémantiques

## Résolution du *term mismatch*

**Définition** : un terme de la requête est absent dans les documents mais présent dans l'ontologie

- Retrouver de nouveaux documents pertinents : le rappel ↗

REQ : *"I would like a salad with **kiwano**"*



# Perspectives

- Réseaux de neurones
- Apprentissage des poids et valeurs
- Comparaison à d'autres SRI sémantiques : *Open Semantic Search* (https ://www.opensemanticsearch.org/)

| | R-PREC | MAP |
|---|---|---|
| $BM25$ | 0.30 | 0.30 |
| $TDintT$ | 0.18 | 0.18 |
| $BM25$ (documents courts) | 0.31 | 0.32 |
| $TDintT$ (documents courts) | 0.22 | 0.23 |

| | R-PREC | MAP |
|---|---|---|
| $TDintT$ | 0.18 | 0.18 |
| $TDCintT_{Manu}$ | 0.19 | 0.19 |

| | R-PREC | MAP |
|---|---|---|
| $TDCintT_{Manu}$ | 0.19 | 0.19 |
| $TDCintC_{Manu}$ | 0.11 | 0.12 |
| $TDCintTC_{Manu}$ | 0.21 | 0.20 |

| | R-PREC | MAP |
|---|---|---|
| $TDCintT_{MM}$ | 0.18 | 0.19 |
| $TDCintTC_{Manu}$ | 0.21 | 0.20 |
| $TDCintTC_{MM}$ | **0.30** | **0.30** |

| | R-PREC | MAP |
|---|---|---|
| $TDintT$ | 0.18 | 0.18 |
| $TDintT_{TSTC}$ | 0.21 | 0.21 |
| $DTCintT_{MM}$ | 0.22 | 0.22 |
| $DTCintTC_{MM}$ | 0.21 | 0.22 |

| | R-PREC | MAP |
|---|---|---|
| $TDintT_{TSTC}$ | 0.21 | 0.21 |
| $DTCintT_{MM}$ | 0.22 | 0.22 |
| $DTCintTC_{MM}$ | 0.21 | 0.22 |
| $DTCDintT_{MM}$ | 0.23 | 0.24 |
| $DTCDintTC_{MM}$ | 0.26 | 0.28 |

- **Taille du graphe et temps d'exécution**

| Modèle | Taille du graphe | | Temps | | NB it |
|---|---|---|---|---|---|
| | $N$ | $R$ | Const $G$ | PA sur $G$ | |
| $TD$ | 106 184 | 5 029 046 | 2h 25min | 1jr 38min | 6 |
| $DTC$ | 165 554 | 5 829 516 | 9h 54min | 1jr 6h 32min | 8 |
| $DTCD$ | 165 554 | 7 772 340 | 11h 17min | 1jr 17h 15min | 6 |

Aronson, A. R. (2001).
Effective mapping of biomedical text to the umls metathesaurus : the metamap program.
Proc AMIA Symp, pages 17–21.

Bannour, I. and Zargayouna, H. (2012).
Une plate-forme open-source de recherche d'information sémantique.
In COnférence en Recherche d'Information et Applications (CORIA), pages 167–178.

Baziz, M. (2005).
Indexation conceptuelle guidée par ontologie pour la recherche d'information.
PhD thesis.

Bhagdev, R., Chapman, S., Ciravegna, F., Lanfranchi, V., and Petrelli, D. (2008).
Hybrid search : effectively combining keywords and semantic searches.

In Proceedings of the 5th European semantic web conference on The semantic web : research and applications, ESWC'08, pages 554–568, Berlin, Heidelberg. Springer-Verlag.

Bhogal, J., Macfarlane, A., and Smith, P. (2007).
A review of ontology based query expansion.
Information Processing and Management, 43(4) :866 – 886.

Bikakis, N., Giannopoulos, G., Dalamagas, T., and Sellis, T. (2010).
Integrating keywords and semantics on document annotation and search.
In Proceedings of the 2010 international conference on On the move to meaningful internet systems : Part II, OTM'10, pages 921–938, Berlin, Heidelberg. Springer-Verlag.

Boubekeur, F. and Azzoug, W. (2013).
Concept-based indexing in text information retrieval.
CoRR, abs/1303.1703.

Cohen, P. R. and Kjeldsen, R. (1987).

Information retrieval by constrained spreading activation in semantic networks.
Inf. Process. Manage., 23(4) :255–268.

Crestani, F. (1997).
Application of spreading activation techniques in information retrieval.
Artificial Intelligence Review, 11(6) :453–482.

Croft, W., Lucia, T., Cringean, J., and Willett, P. (1989).
Retrieving documents by plausible inference : An experimental study.
Information Processing and Management, 25(6) :599 – 614.

Dinh, D. and Tamine, L. (2010).
Vers un modèle d'indexation sémantique adapté aux dossiers médicaux de patients.
Conférence francophone en Recherche d'Information et Applications, CORIA 2010, pages 325–336.

Fernandez, M., Cantador, I., Lopez, V., Vallet, D., Castells, P., and Motta, E. (2011).

Semantically enhanced information retrieval : an ontology-based approach.
Web Semantics : Science, Services and Agents on the World Wide Web, 9(4).

Gerani, S., Zhai, C., and Crestani, F. (2012).
Score transformation in linear combination for multi-criteria relevance ranking.
In Advances in Information Retrieval - 34th European Conference on IR Research, ECIR 2012, Barcelona, Spain, April 1-5, 2012. Proceedings, pages 256–267.

Gonzalo, J., Verdejo, F., Chugur, I., and Cigarrán, J. M. (1998).
Indexing with wordnet synsets can improve text retrieval.
CoRR, cmp-lg/9808002.

Hamadan, H., Albitar, S., Bellot, P., Espinasse, B., and Fournier, S. (2012).
Lsis at trec 2012 medical track – experiments with conceptualization, a dfr model and a semantic measure.

In The Twenty-First Text REtrieval Conference (TREC 2012) Notebook, volume Special Publication, page 12 p., Gaithersburg (USA).

Jiang, X. and Tan, A.-H. (2006).
Ontosearch : A full-text search engine for the semantic web.
In Proceedings of the 21st National Conference on Artificial Intelligence - Volume 2, AAAI'06, pages 1325–1330. AAAI Press.

Khan, L. R. (2000).
Ontology-based Information Selection.
PhD thesis, Faculty of the Graduate School, University of Southern California.

Koopman, B., Zuccon, G., Nguyen, A., Vickers, D., Butt, L., and Bruza, P. D. (2012).
Exploiting snomed ct concepts and relationships for clinical information retrieval : Australian e-health research centre and queensland university of technology at the trec 2012 medical track.

In 21st Text REtrieval Conference (TREC 2012), pages 1–8, Gaithersburg, Md. National Institute of Standards and Technology - NIST.

Preece, S. (1981).
A Spreading Activation Network Model for Information Retrieval.
University of Illinois at Urbana-Champaign.

Rocha, C., Schwabe, D., and Aragao, M. P. (2004).
A hybrid approach for searching in the semantic web.
In Proceedings of the 13th International Conference on World Wide Web, WWW '04, pages 374–383, New York, NY, USA. ACM.

Salton, G. (1968).
Automatic Information Organization and Retrieval.
McGraw Hill Text.

Salton, G. and Buckley, C. (1988).
On the use of spreading activation methods in automatic information.

In SIGIR '88 : Proceedings of the 11th annual international ACM SIGIR conference on Research and development in information retrieval, pages 147–160, New York, NY, USA. ACM Press.

Savoy, J. (1992).
Bayesian inference networks and spreading activation in hypertext systems.
Information Processing and Management, 28(3) :389 – 406.

Schumacher, K., Sintek, M., and Sauermann, L. (2008).
Combining fact and document retrieval with spreading activation for semantic desktop search.
In The Semantic Web : Research and Applications, volume 5021 of Lecture Notes in Computer Science, pages 569–583. Springer Berlin Heidelberg.

Shoval, P. (1981).
Expert/consultation system for a retrieval data-base with semantic network of concepts.

In Proceedings of the 4th Annual International ACM SIGIR Conference on Information Storage and Retrieval : Theoretical Issues in Information Retrieval, SIGIR '81, pages 145–149, New York, NY, USA. ACM.

Stetina, J., Kurohashi, S., and Nagao, M. (1998).
General word sense disambiguation method based on a full sentential context.
In IN USAGE OF WORDNET IN NATURAL LANGUAGE PROCESSING, PROCEEDINGS OF COLING-ACL WORKSHOP.

Wang, H., Tran, T., Liu, C., and Fu, L. (2011).
Lightweight integration of ir and db for scalable hybrid search with integrated ranking support.
Web Semant., 9(4) :490–503.

Zargayouna, H. (2005).
Indexation sémantique de documents XML.
PhD thesis, Université Paris-Sud.

Zhang, L., Liu, Q., Zhang, J., Wang, H., Pan, Y., and Yu, Y. (2007).

Semplore : An ir approach to scalable hybrid query of semantic web data.
In Aberer, K., Choi, K.-S., Noy, N., Allemang, D., Lee, K.-I., Nixon, L. J. B., Golbeck, J., Mika, P., Maynard, D., Schreiber, G., and Cudré-Mauroux, P., editors, Proceedings of the 6th International Semantic Web Conference and 2nd Asian Semantic Web Conference (ISWC/ASWC2007), Busan, South Korea, volume 4825 of LNCS, pages 645–658, Berlin, Heidelberg. Springer Verlag.