

# Vers la recherche d'informations sur le Web basée sur des données d'opinion

Mehdi Adda : doctorant Université de Montréal/Université de Lille

Rokia Missaoui : Prof. Université du Québec en Outaouais

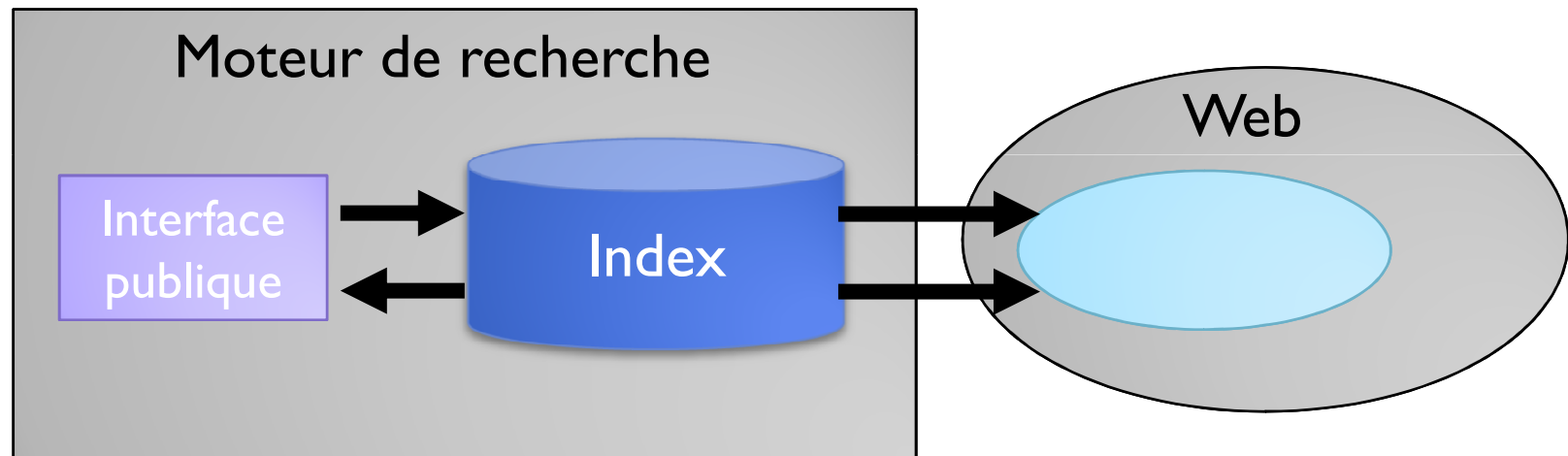
Petko Valtchev : Prof. Université du Québec à Montréal



# Plan

- Principes de base d'un moteur de recherche
- Données d'opinion pour la recherche Web
- Proposition
- Modélisation et implémentation
- Conclusion et perspectives

# Moteur de recherche Web





ruby java

Search

Advanced Search  
Preferences

Search:  the web  pages from Canada

Web

Results 1 - 10 of about 3,000,000 for [ruby java](#).

### [Ruby for the Java world - Java World](#)

17 Jul 2006 ... Dynamic languages are the up-and-coming competitors to **Java**. Expanding their scope beyond glue code and Web GUIs into the heavy-duty ...

[www.javaworld.com](#)

### [Using Ruby](#)

Small tutorial w  
[onestepback.org](#)

### [ONJava.com](#)

16 Nov 2005 ...

**Java**. Tate, Jan

[www.onjava.com](#)

### [To Ruby From](#)

And it's fast (co

Going from **Java** to **Ruby**, expect your code size to shrink down ...

[www.ruby-lang.org/en/documentation/ruby-from-other-languages/to-ruby-from-java/](#) - 10k -

[Cached](#) - [Similar pages](#)

### [Ruby Java Bridge](#)

You can bind **Ruby** object to **Java** interface as long as the object has responsible to respond the method call from **Java** world. ...

[rjb.rubyforge.org/](#) - 12k - [Cached](#) - [Similar pages](#)

Sponsored Links

### [Ruby Enterprs.Integration](#)


[Ruby on J2EE servers: JRuby](#)  
[Integration/LegacyApps](#)  
[e.ca](#)

- **Constat :**
  - Un mot clé => plusieurs ressources
- **Question :**
  - Comment classer ces ressources ?



# I<sup>ère</sup> génération : "dans la page"

- La localisation et la fréquence des mots clés dans la page influence l'ordre de classement
- Inconvénient
  - Choix des mots clés et de leur emplacement ... *pollupostage*



## 2<sup>ème</sup> génération : "à l'extérieur de la page"

- **Utiliser des facteurs externes à la page**
- Difficiles à être manipulées par les webmasters
  - Total de clics
  - Liens externes
  - Mots contenus dans les liens
- **Abus**
  - Échange/achat/vente de liens

# Limitation

- Classement automatique des pages
  - Degré de pertinence calculé d'une page <> degré de pertinence effectif
- Questions
  - Où chercher la pertinence effective ?
  - Comment exploiter cette pertinence ?



# Recherche d'informations basée sur les données d'opinion

- Pourquoi ?
  - Améliorer la recherche d'informations sur les web
- Comment ?
  - Prendre en compte les données d'opinion déjà disponibles sur des sites d'évaluation de contenu Web
  - La pertinence effective est approchée par les évaluations des utilisateurs



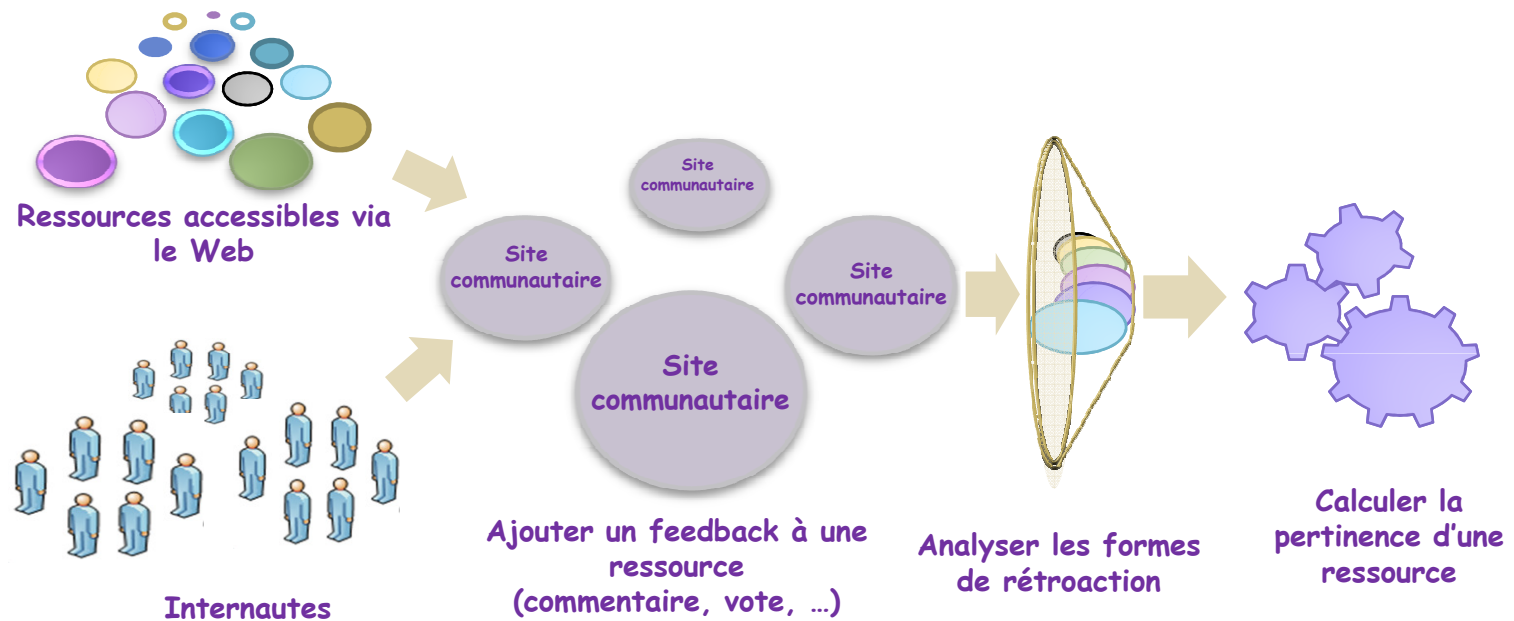
# Moteurs de recherches basés sur les évaluations des utilisateurs

- **Exemples**
  - Mahalo, Dipiti
- **Inconvénients**
  - Bases d'opinion propres aux moteurs
  - Non exploitation des évaluations déjà disponibles

# Proposition

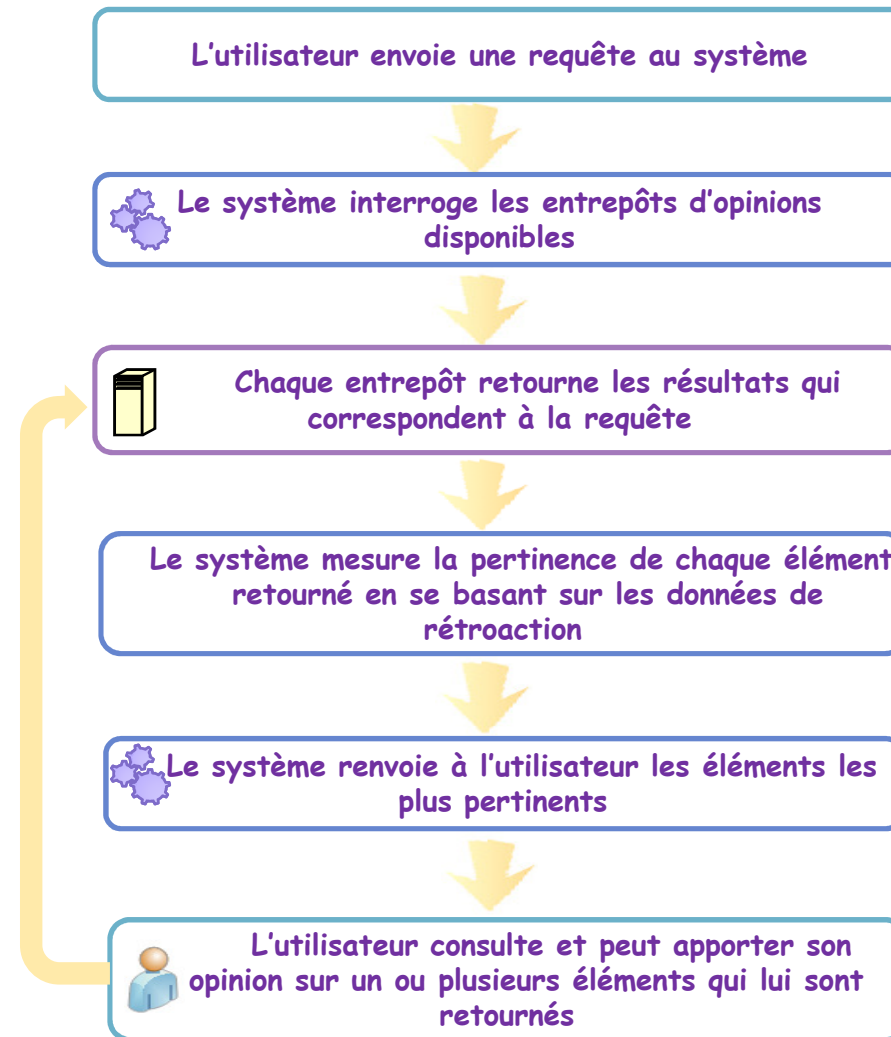
- Recherche basée sur la combinaison des
  - critères classiques de recherche (mots clés)
  - évaluations des internautes **laissées sur des sites existants**
- Exemples de sites d'évaluation de contenu
  - Slashdot, Reddit, Dzone, Digg, ...

# Processus d'intégration des données d'opinion dans la recherche Web



Site communautaire d'évaluation de ressources : entrepôt d'opinion

# Scénario type





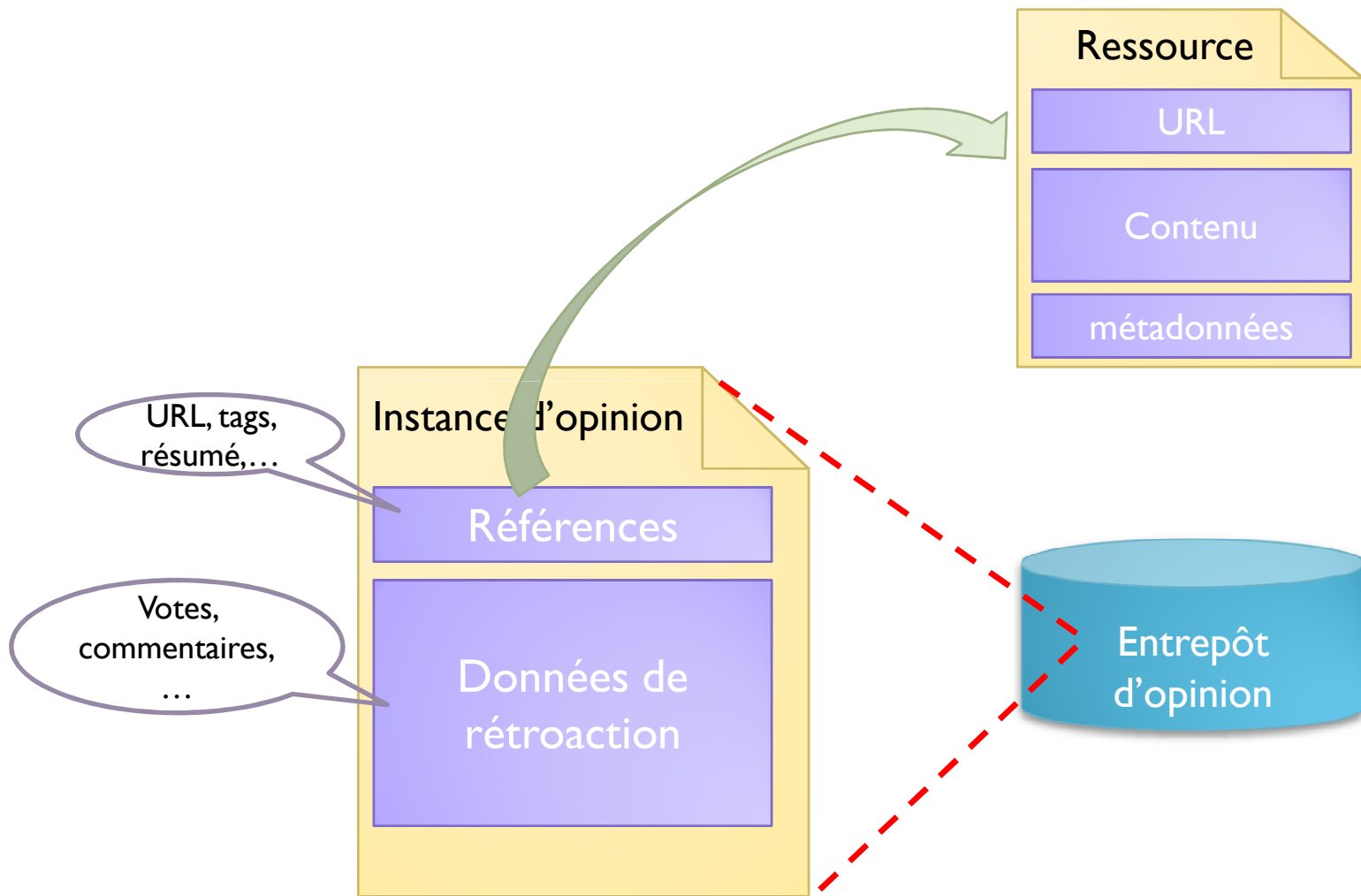
Comment réaliser un tel système ?



# Modélisation

- Description des données d'opinion
- Langage de description de classes de rétroaction
  - Syntaxe
  - Sémantique
- Ensemble de primitives

# Instance d'opinion

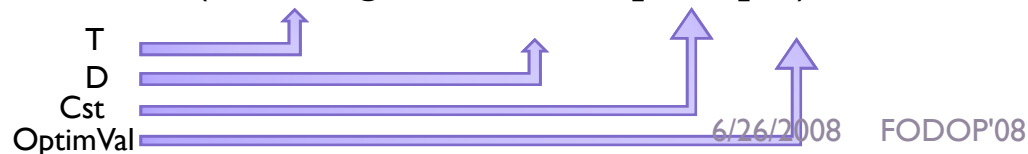


# Classes de rétroaction (I)

- Représenter une catégorie de rétroaction
  - votes positifs, votes négatifs, etc.
- Syntaxe
  - Type de rétroaction (T)
  - Domaine de définition (D)
  - Contraintes (Cst)
  - Valeur optimale (OptimVal)
- Ex.

(VotePositif, Entiers, [0..10], 10)

(VoteNégatifs, Entiers, [0..10], 0)



# Classes de rétroaction (2)

- Sémantique

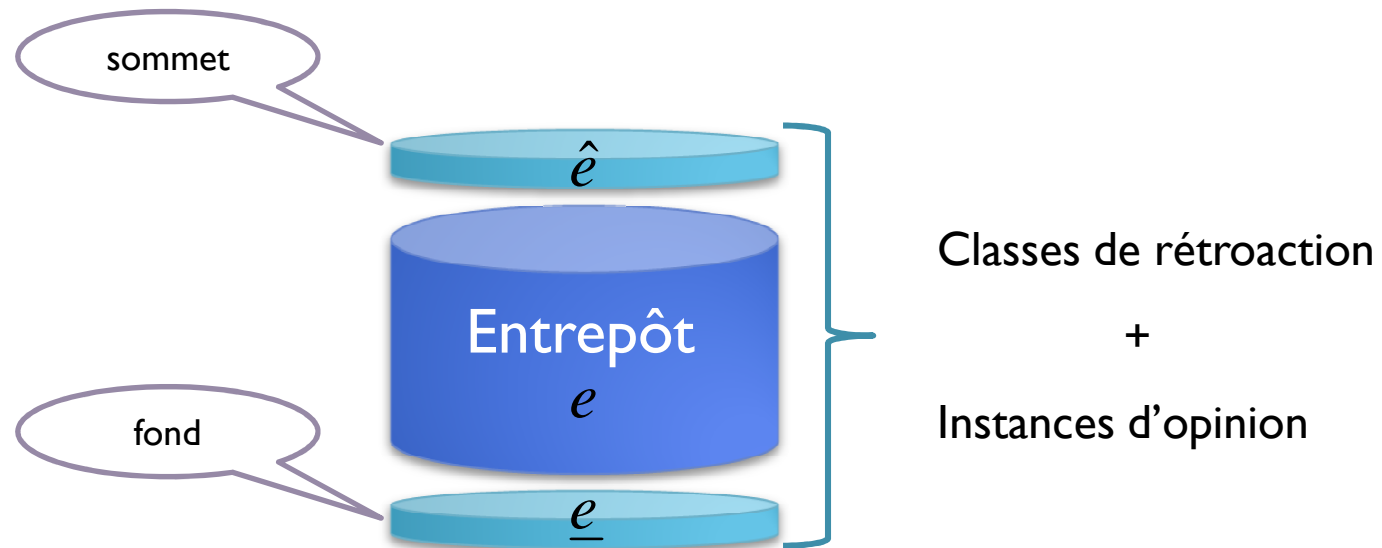
- Relation d'instanciation
  - $R(\text{Classe de rétroaction}, \text{objet de rétroaction})$
- Domaine d'interprétation
  - Objets de rétroaction instances d'une classe

- Ex.

(VotePositif, 5)      **instance de**      (VotePositif, Entiers, [0..10], 10)

(VotePositif, 12) **n'est pas instance de** (VotePositif, Entiers, [0..10], 10)

# Entrepôt d'opinion



Comment calculer la pertinence d'une instance d'opinion ?

# Pertinence d'une instance d'opinion

- S'éloignée du fond  $\Rightarrow$  s'approcher du sommet  
 $\Rightarrow$  **s'approcher des valeurs optimales**
- Distance entre une instance et le fond de l'entrepôt
- Mesure de *Minkowsky* d'ordre  $k$

$$p_e(o) = d_e(o, \underline{e}) = \sqrt[k]{\sum_{i \in [1..l]} (v[i] - \underline{e}[i])^k}$$

- $k$  : nombre de classes d'opinions de  $e$
- $l$  : nombre d'objets de rétroaction de  $o$
- $v[i]$  : valeur du  $i^{\text{ème}}$  objet de rétroaction de  $o$

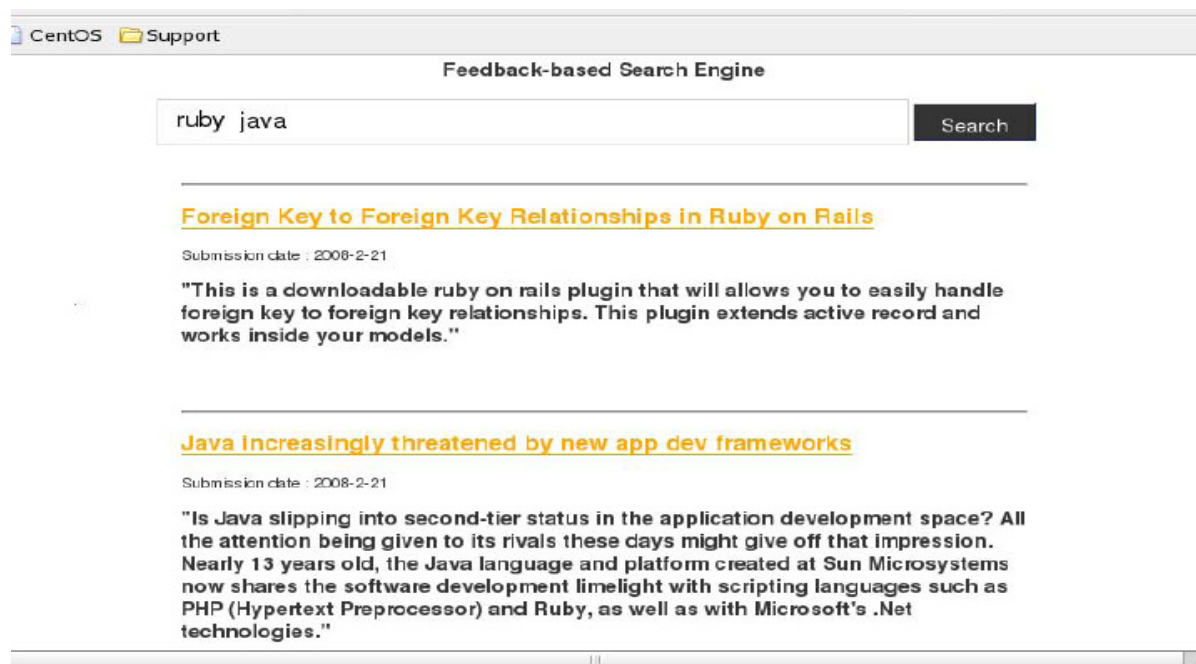
# Normalisation de la mesure de pertinence

- $p()$  n'est pas normalisée
  - Les pertinences de deux instances d'opinion de deux entrepôts différents ne peuvent pas être comparées
- Normalisation

$$\overline{\mathcal{P}_e(o)} = \begin{cases} 0, & \text{si } \mathcal{P}_e(\hat{e}) = 0; \\ |\frac{\mathcal{P}_e(o)}{\mathcal{P}_e(\hat{e})}|, & \text{sinon.} \end{cases}$$

# Implémentation

- *SocialSeeker*
- Paradigme : MVC (Modèle-Vue-Contrôleur)
- Cadre d'application : Ruby On Rails
- Enterpôts : Dzone + Reddit



CentOS Support

Feedback-based Search Engine

ruby java Search

---

**[Foreign Key to Foreign Key Relationships in Ruby on Rails](#)**

Submission date : 2008-2-21

"This is a downloadable ruby on rails plugin that will allows you to easily handle foreign key to foreign key relationships. This plugin extends active record and works inside your models."

---

**[Java increasingly threatened by new app dev frameworks](#)**

Submission date : 2008-2-21

"Is Java slipping into second-tier status in the application development space? All the attention being given to its rivals these days might give off that impression. Nearly 13 years old, the Java language and platform created at Sun Microsystems now shares the software development limelight with scripting languages such as PHP (Hypertext Preprocessor) and Ruby, as well as with Microsoft's .Net technologies."



# Conclusion et perspective

- Cadre d'application pour l'intégration des données d'opinion dans la recherche Web
- Prototype d'un moteur de recherche qui basé sur le cadre proposé
- Données non numériques
- API d'intégration dans les entrepôts d'opinion
- Mise à l'échelle et prise en considération de changements dans les entrepôts de manière asynchrone



# Références

- [ADA 97] ADAH S., BUFI C., TEMTANAPAT Y., « Integrated Search Engine », KDEX '97 : Proceedings of the 1997 IEEE Knowledge and Data Engineering Exchange Workshop, , 1997, p. 140-147.
- [BAA 03] BAADER F., CALVANESE D., MCGUINNESS D., NARDI D., PATEL-SCHNEIDER P., « The Description Logic Handbook : Theory, Implementation and Applications », Cambridge University Press, , 2003, page 574.
- [BOJ 08] BOJRS U., BRESLIN J. G., FINN A., DECKER S., « Using the Semantic Web for linking and reusing data across Web 2.0 communities », Web Semant., vol. 6, no 1, 2008, p. 21-28, Elsevier Science Publishers B. V.
- [BOR 03] BORGATTI S.-P., CROSS R., « A Relational View of Information Seeking and Learning in Social Networks », Manage. Sci., vol. 49, no 4, 2003, p. 432-445, INFORMS.
- [DIN 03] DING C., CHI C.-H., « A Generalized Site Ranking Model for Web IR », , 2003, p. 584-587, IEEE Computer Society.



Merci pour votre attention

Questions ?