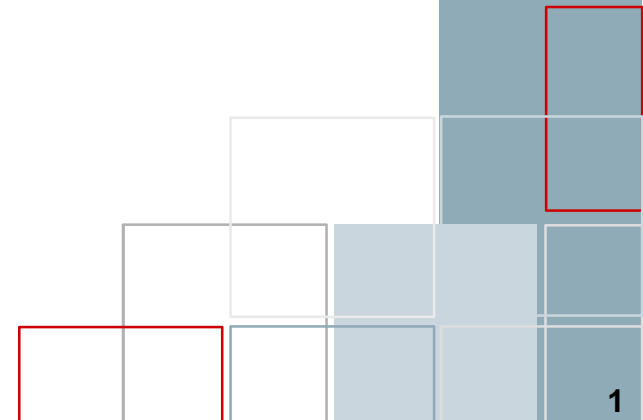


# **Veille technologique en télécommunications**

**Khalid EL HIMDI**  
**Ismail KASSOU**  
**Hamid MACHCHOUR**





- L'analyse de contenus est au cœur de la gestion du cycle de vie de l'information textuelle.
- Ces techniques permettent de manipuler intelligemment les documents écrits.
- Il existe deux grands types de techniques d'analyse de contenu :
  - Le **traitement automatique du langage** (TAL) pour l'analyse fine d'un nombre peu important de documents
  - Le **textmining** pour traiter de gros volumes de données pour en retirer des informations non visibles.

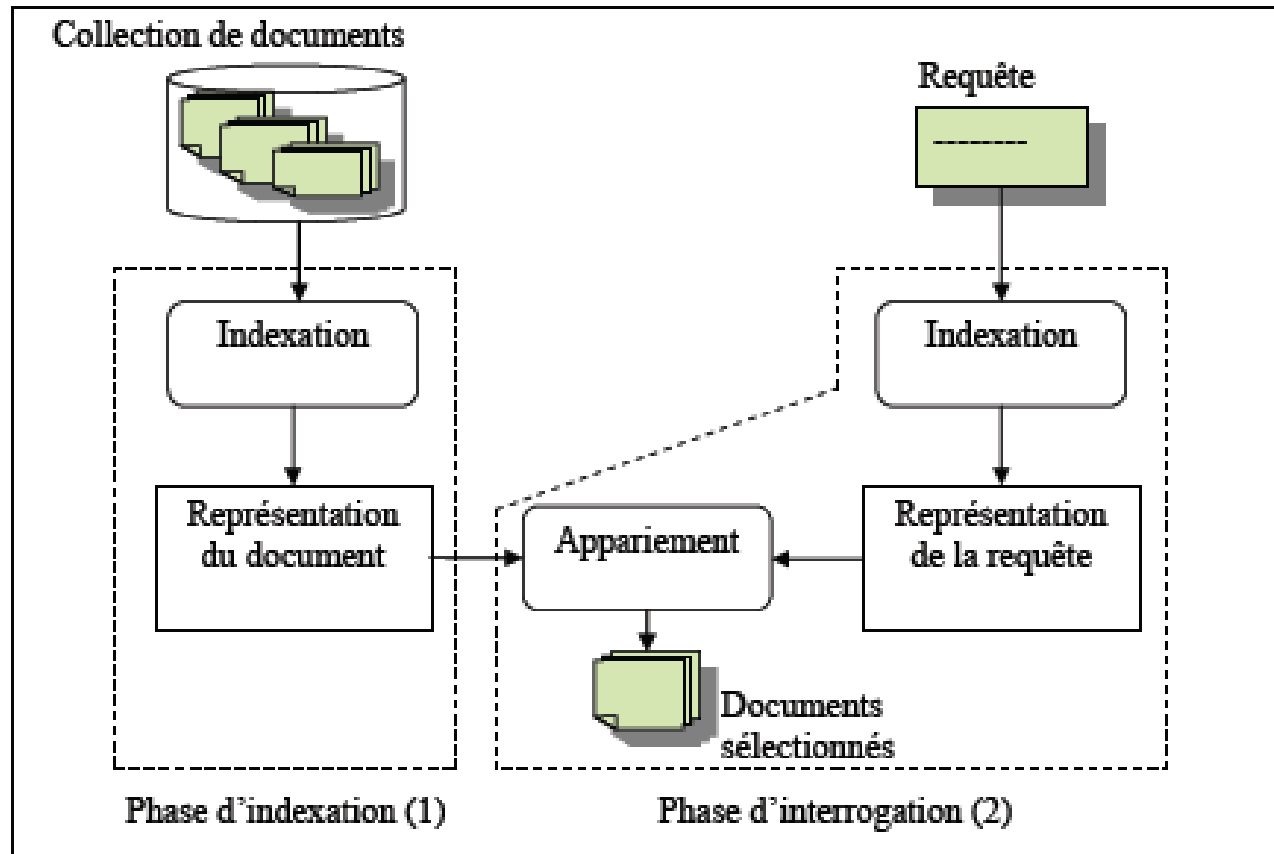


- Les systèmes de TAL reposent sur des bases de connaissance.
- Une base de connaissance peut contenir les informations morpho-lexicales (lexique, dictionnaire), syntaxiques, sémantiques (thésaurus) et parfois mêmes conceptuelles (réseaux sémantiques).
- On recense cinq ressources linguistiques utilisées :
  - Les lexiques,
  - Les dictionnaires,
  - Les grammaires,
  - Les thésaurus,
  - Les réseaux sémantiques.



# Recherche de documents

ENSIAS





- Consécutivement au traitement d'une requête par le système, les documents de la collection forment deux partitions selon deux caractéristiques :
  - les documents non pertinents non restitués
  - les documents non pertinents restitués (bruit )
  - les documents pertinents restitués
  - les documents pertinents non restitués (silence)



# Amélioration des moteurs de recherche

ENSIAS

- Les services d'extensions sémantiques des moteurs de recherche dépendent de la richesse des ressources terminologiques qui leur sont fournies et de leur capacité à les utiliser.
- Pour améliorer le fonctionnement des moteurs de recherche nous devons leur fournir :
  - Un vocabulaire issu de la terminologie
  - Une organisation hiérarchique des termes
  - Des liens de connaissances, relations sémantiques



# Amélioration des moteurs de recherche

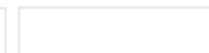
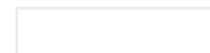
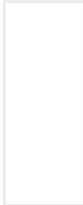
ENSIAS

- L'extension des capacités et l'amélioration de l'efficacité des moteurs de recherche par l'utilisation de terminologies métier.
- Les raisons de ces besoins peuvent avoir deux origines cumulatives :
  - L'information que nous cherchons n'est pas en ligne et nous n'avons donc aucune chance de la trouver mais également aucune chance d'être certain qu'elle n'est pas en ligne
  - Nous ne savons pas comment décrire notre problématique pour avoir une chance de trouver le document utile, lui même rédigé avec le vocabulaire métier/jargon de l'entreprise ou de l'administration



# Approche proposée pour la veille en télécommunications

ENSIAS







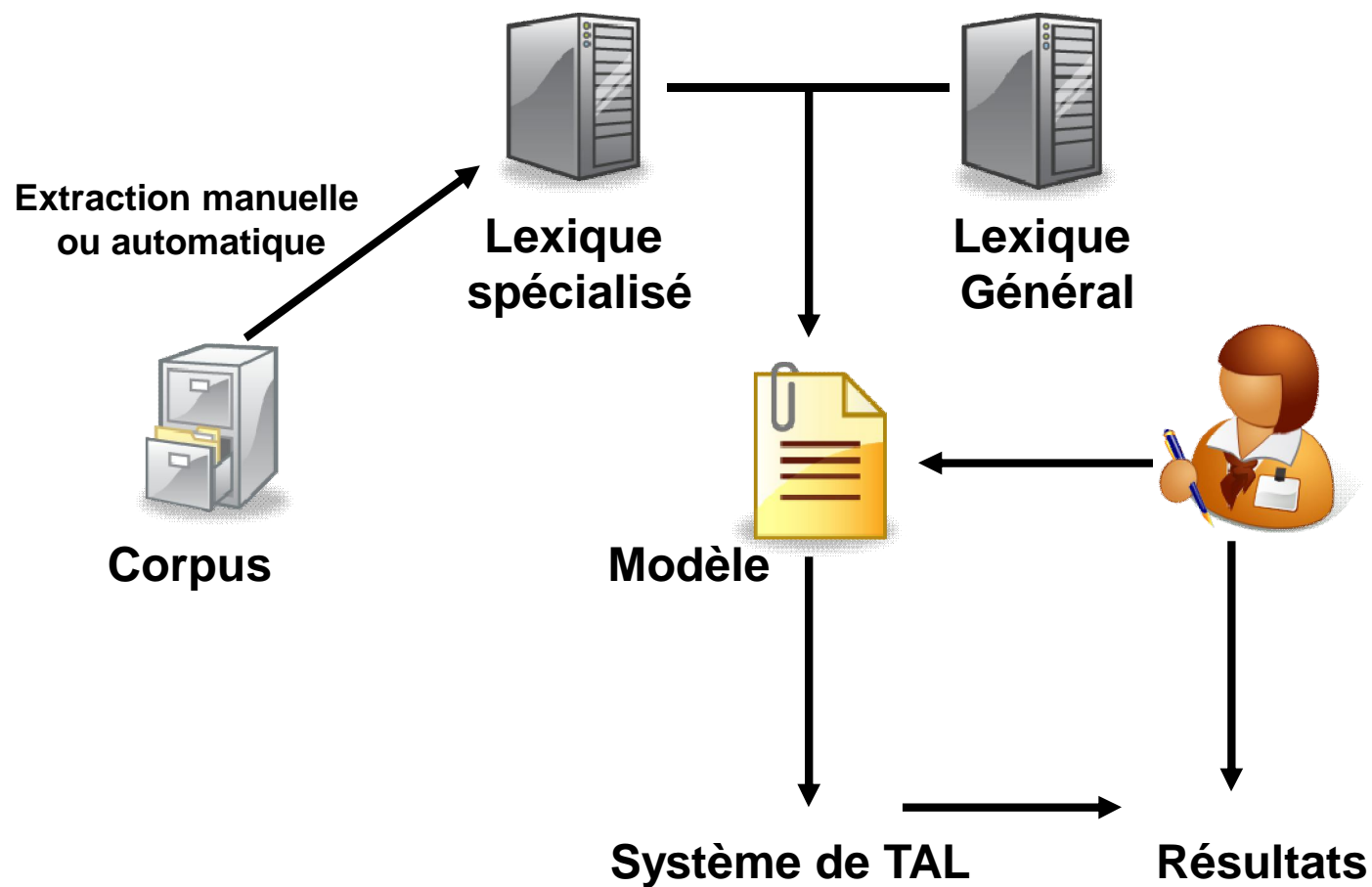
# Présentation générale

ENSIAS

- Dans le cadre d'un projet R&D avec Maroc Telecom, nous nous sommes intéressés au développement de techniques destinées à l'amélioration de la recherche de documents pertinents sur le Web.
- Nous proposons une solution basée sur l'exploitation d'ontologies, fondements du Web Sémantique, pour représenter les connaissances du domaine de recherche visé par des requêtes ainsi que les vues des utilisateurs sur ce domaine.
- Les ontologies sont vues comme des connaissances qui évoluent au cours du temps.



# Approche par corpus





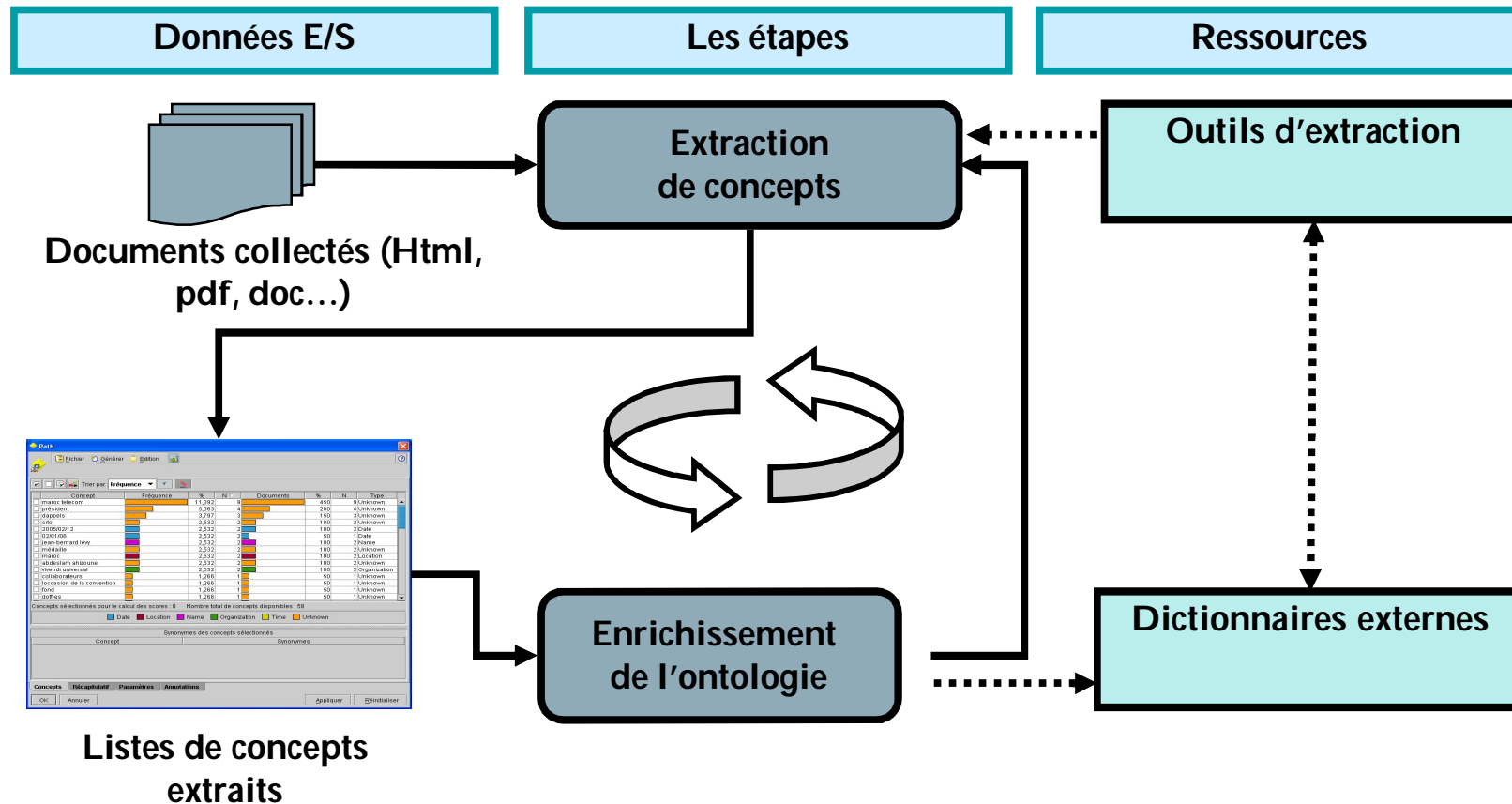
# Constitution du lexique

ENSIAS

- C'est une étape critique car elle détermine le bon fonctionnement de l'application.
- Il faut utiliser au mieux la spécificité linguistique du domaine (abréviations, expressions, mode de rédaction).
- On peut utiliser des méthodes statistiques capables d'extraire des phrases ou des mots représentatifs d'un corpus.



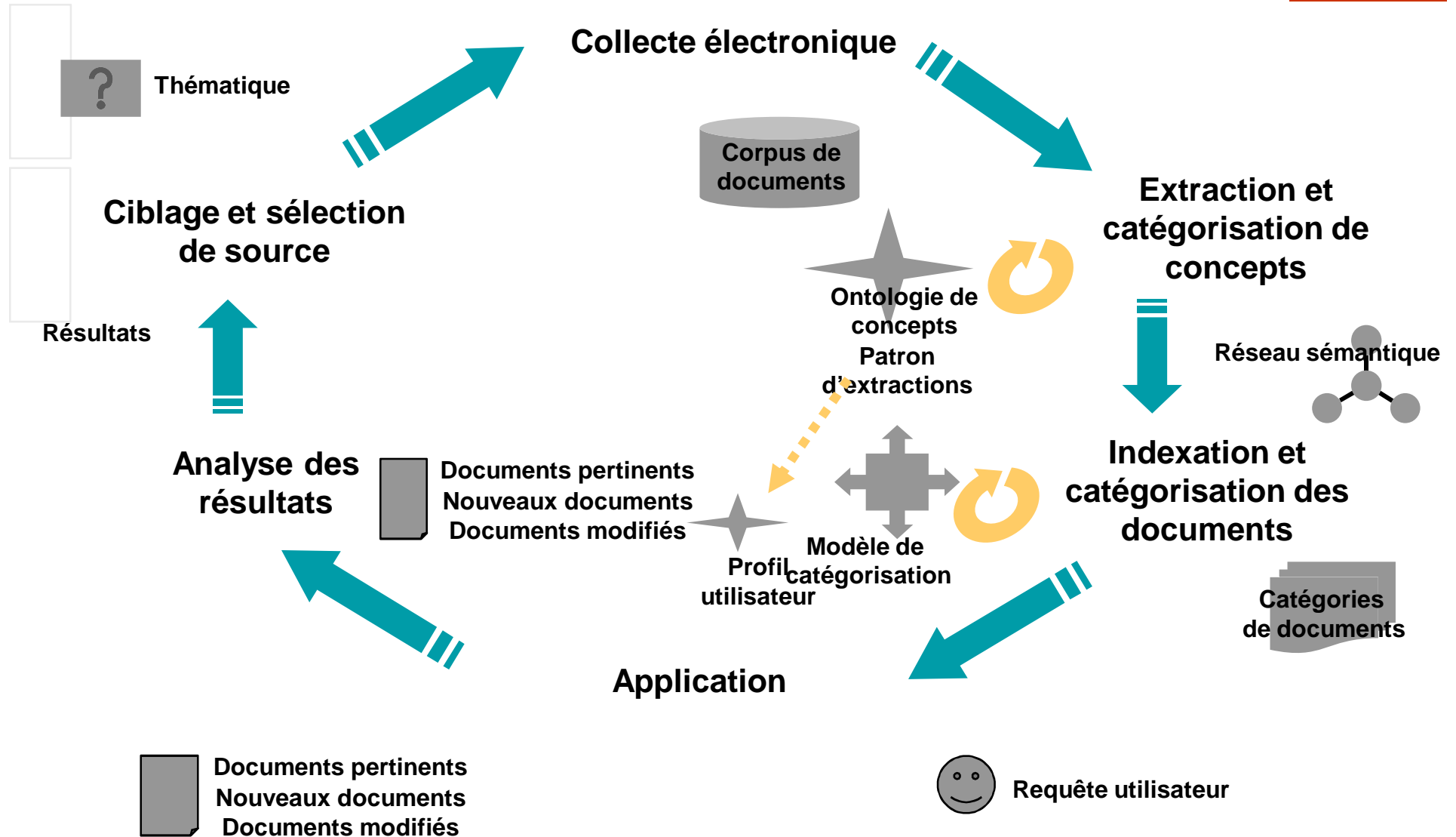
# Approche adoptée pour l'extraction et l'affinement de l'ontologie





# Processus global de veille

ENSIAS





# Notre approche pour la veille

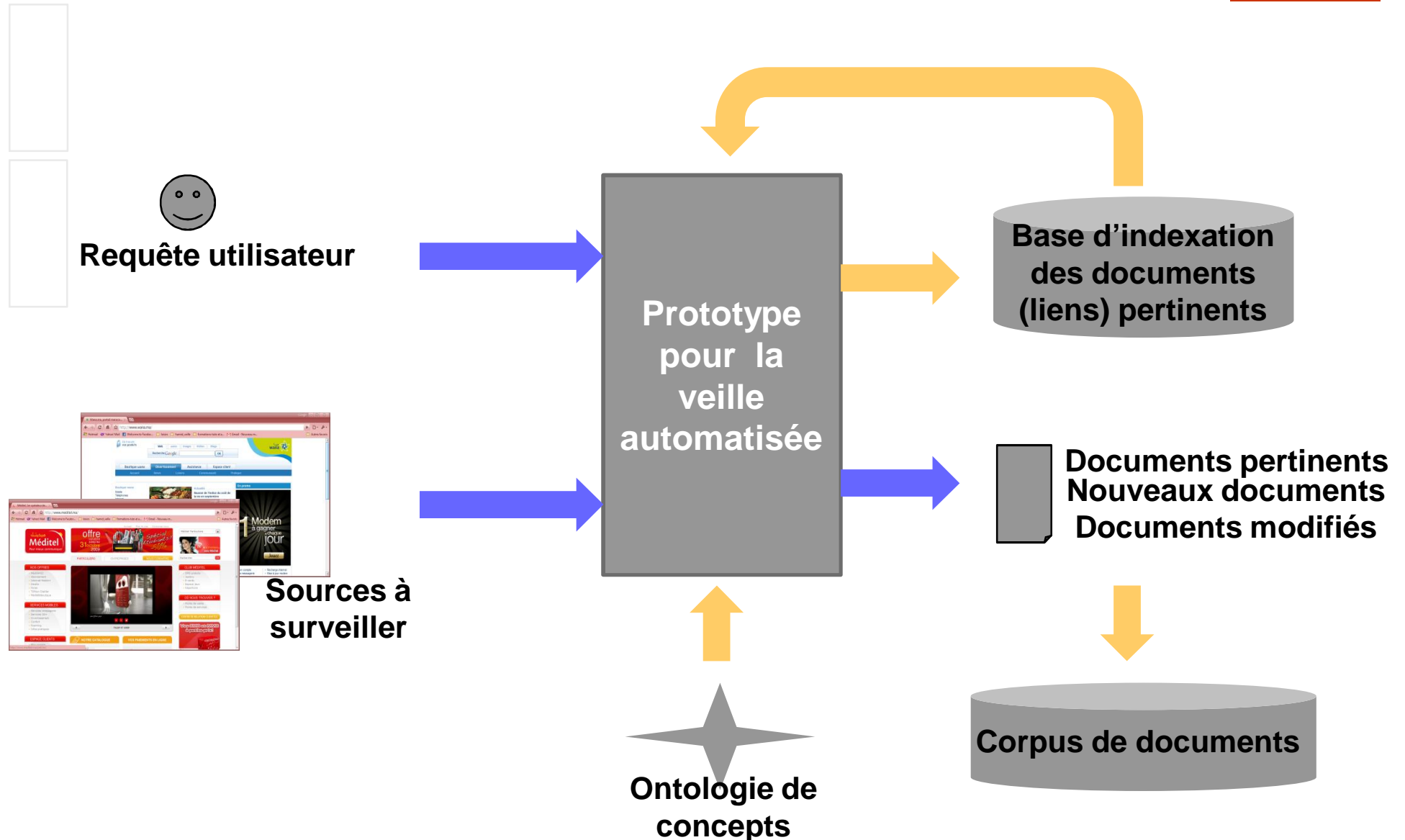
ENSIAS

- Approche semi-automatique,
- Basé sur la construction d'une ontologie du domaine recherché grâce à un corpus tiré du web et comprenant des documents types,
- Basé sur la mise en place de profils utilisateurs basés sur l'ontologie développée,
- Basé sur une recherche en partie indépendante des moteurs de recherche.



# Prototype de l'outil de surveillance automatisée des sources cibles

ENSIAS





# Exemple de résultats

SCSCCorpus - Prototype version

Configuration | Contrôle | **Resultats**

Anciennes URLs: 0

URLs modifiées: 0

Nouvelles URLs: 36

URL	PERTINENCE	TERMES TROUVES
<a href="http://www.iam.ma/Entreprises/Services/Recharge-EI-Manzil.aspx">http://www.iam.ma/Entreprises/Services/Recharge-EI-Manzil.aspx</a>	0.5238894928500966	terminaux(1.0),télép...
<a href="http://www.iam.ma/Particuliers/Fixe/SERVICES/Pages/Recharge-EI-Manzil.aspx">http://www.iam.ma/Particuliers/Fixe/SERVICES/Pages/Recharge-EI-Manzil.aspx</a>	0.517431610873925	carte prépayée(1.0),t...
<a href="http://www.inwi.ma/mot-clés">http://www.inwi.ma/mot-clés</a>	0.4742688878808664	téléphone(4.0),
<a href="http://www.iam.ma/Groupe">http://www.iam.ma/Groupe</a>	0.4742688878808664	terminaux(1.0),
<a href="http://www.inwi.ma/mot-clés">http://www.inwi.ma/mot-clés</a>	0.4742688878808664	téléphone(1.0),
<a href="http://www.inwi.ma/entree">http://www.inwi.ma/entree</a>	0.4742688878808664	téléphone(2.0),
<a href="http://www.inwi.ma/mot-clés">http://www.inwi.ma/mot-clés</a>	0.4742688878808664	téléphone(1.0),
<a href="http://www.inwi.ma/corpus">http://www.inwi.ma/corpus</a>	0.4742688878808664	téléphone(1.0),
<a href="http://www.inwi.ma/mot-clés">http://www.inwi.ma/mot-clés</a>	0.4742688878808664	téléphone(2.0),

URL:

TITRE:

CONCEPTS TROUVES:

MOTS CLES:

DESCRIPTION:

CONTENU:

DATE:  TAILLE(Octet):





# Exemple de résultats

ENSIAS

SCSCCorpus - Prototype version

Configuration Contrôle **Resultats**

**Anciennes URLs: 0**

**URLs modifiées: 0**

**Nouvelles URLs: 7**

URL	PERTINENCE	TERMES TROUVES
<a href="http://www.iam.ma/Parti">http://www.iam.ma/Parti</a>	0.6173165676349103	jawal(1.0),
<a href="http://www.iam.ma/Entre">http://www.iam.ma/Entre</a>	0.6173165676349103	jawal(2.0),
<a href="http://www.iam.ma/Entre">http://www.iam.ma/Entre</a>	0.6173165676349103	jawal(5.0),
<a href="http://www.inwi.ma/mob">http://www.inwi.ma/mob</a>	0.6173165676349103	tictac(4.0),
<a href="http://www.iam.ma/cont">http://www.iam.ma/cont</a>	0.6173165676349103	jawal(1.0),
<a href="http://www.inwi.ma/corp">http://www.inwi.ma/corp</a>	0.6173165676349103	tictac(1.0),
<a href="http://www.iam.ma/Parti">http://www.iam.ma/Parti</a>	0.6173165676349103	jawal(1.0),

**URL:**

**TITRE:**

**CONCEPTS TROUVES:**

**CONTENU:**  
... et avec tictac de inwi ...  
... combien coûtent les appels tictac ...  
... ce service est disponible par défaut dans votre offre tictac ...  
... avec tictac ...

**MOTS CLES:**

**DESCRIPTION:**

**DATE:**

**TAILLE(Octet):**



**Merci de votre attention**



# Remerciements

ENSIAS

Nous tenons à remercier Maroc Télécom pour avoir accepté de financer ce projet de recherche et pour son aide afin de mener à terme ce projet.