

## Panorama des outils de veille

Joumana Boustany

Maître de conférences  
Université Paris 5 – René Descartes  
Institut Universitaire de Technologie

19e journée RNDH - 21/09/2006

## Aujourd'hui

- Information abondante et diversifiée
- Le domaine médical
  - vocabulaire spécifique
  - sources informelles limitées malgré un développement des sources purement Web : portails, newsletters, fils RSS...
  - **Le Web invisible (bases de données structurées : articles de presse, brevets, thèses...) reste incontournable**
- Le domaine administratif français et européen
  - Administration très présente sur le Web
- Chaque domaine ses spécificités

19e journée RNDH - 21/09/2006

## Les agents intelligents

- Automatisation des requêtes récurrentes
- Autonomie
- Auto-apprentissage

**Actuellement peu d'agents intelligents répondent à ces différents critères**

19e journée RNDH - 21/09/2006

## Panorama des outils (1)

- Outils d'aide au sourcing (social bookmarking)
  - <http://del.icio.us/>
- Outils de collecte :
  - Agents de recherche ([Copernic Agent](#))
  - Agents de surveillance ([KB-Crawl](#), [Website watcher](#), [Copernic Tracker](#)...)
  - *Aspirateurs de sites (MemoWeb)*

19e journée RNDH - 21/09/2006

## Panorama des outils (2)

- Outils d'analyse et de traitement de l'information
  - Techniques linguistiques (calcul des occurrences)
  - Syntaxiques et morphosyntaxiques (décomposition du texte et lemmatisation)
  - Sémantiques (associations de sens)
- Outils de résumé de l'information
  - [Copernic summerizer](#)
  - [Pertinence summarizer](#)

19e journée RNDH - 21/09/2006

## Panorama des outils (3)

- Outils de classification et catégorisation
  - Reconnaissance et identification de concepts
  - Génération automatique de plans de classement (a posteriori)
  - Catégorisation selon un plan préétabli (a priori)
- Outils d'extraction de connaissances (data ou text mining)
  - Permettent d'identifier les concepts et d'établir des relations entre ces éléments

19e journée RNDH - 21/09/2006

## Panorama des outils (4)

- Outils de cartographie
  - Pour visualiser les résultats d'un corpus et faire émerger des liens, des réseaux
- Solutions intégrées qui traitent de tous les processus de veille :
  - collecte multi-sources et multi-bases
  - surveillance
  - catégorisation automatique
  - publication

**Les frontières entre ces différents types d'outils ne sont pas bien délimitées**

19e journée RNDH - 21/09/2006

## Quelques solutions

- [Arisem](#), [Autonomy](#), [Intuition](#), [Go Albert](#)
- [LexiQuest](#), [Lingway](#), [Temis](#)
- etc.

**Ces outils sont payants**

**Pour choisir :**

**Faire une analyse approfondie des besoins**

19e journée RNDH - 21/09/2006

## Faire de la veille à moindre coût

- Les moteurs de recherche ([Exalead](#), [Google](#), [Windows Live Search](#), [Yahoo](#), etc.)
- Les métachercheurs (Copernic, [Vivisimo](#), etc.)
- Les agents d'alerte
- Les desktops
- Les outils spécialisés
- Les blogs et les fils RSS

19e journée RNDH - 21/09/2006

## Pratiques et usages

- En novembre 2002, 48% des internautes s'arrêtaient à la première page de résultats
- En avril 2006, ils seraient 62%, soit une augmentation de 14% !
- Seuls 10% des internautes vont au-delà de la 3<sup>e</sup> page alors qu'ils étaient 19% il y a quatre ans...

Source : [Jupiter Research / iProspect](#)

19e journée RNDH - 21/09/2006

## Google (1)

- [Or blanc](#) => Google ne verra que du blanc
- [Livret A](#) => où est le A ?
- [Jeux de dés](#) => mais où sont les dés ???
- [evenement](#) = [événement](#) = [èvenement](#) ?
- [publicité](#) ≠ [publicités](#) ?

19e journée RNDH - 21/09/2006

## Google (2)

- Le titre
  - [intitle:drogue](#)
  - [intitle:drogue jeunes](#)
  - [allintitle: drogue jeunes](#)
- L'URL
  - [inurl:drogue](#)
  - [inurl:drogue jeunes](#)
  - [allinurl: drogue jeunes](#)

19e journée RNDH - 21/09/2006

## Google (3)

- Un diaporama
  - [filetype:ppt dépenses santé France](filetype:ppt:dépenses%20santé%20France)
- Un document Word
  - [drogue jeunes filetype:doc](drogue%20jeunes filetype:doc)
- Un domaine et ou nom de domaine
  - <site:gouv.fr> "santé publique"
  - <site:europa.eu> santé

19e journée RNDH - 21/09/2006

## Exalead

- Moteur de recherche francophone
- Pas la même puissance que Google, mais des fonctionnalités spécifiques
  - Approximation phonétique
  - Lemmatisation
  - Troncature
  - Clustering : se fait par analyse statistique des termes les plus fréquemment associés à la requête
- Phase de transition => [nouvelle version](#)

19e journée RNDH - 21/09/2006

## Les métachercheurs clients

- Recherche simultanée sur plusieurs outils de recherche (annuaires, moteurs, portails, bases de données...)
- Paramétrage plus avancé que sur les métachercheurs en ligne
- **Copernic Agent** :
  - vérification des liens,
  - téléchargement des résultats,
  - la version payante propose :
    - des fonctions de veille sur les pages et sur les mots-clés
    - des résumés automatiques

19e journée RNDH - 21/09/2006

## Les agents d'alerte

- Signalent les modifications à l'intérieur d'une page
  - Agents d'alerte en ligne  
Infominder <http://www.infominder.com>
- Agents d'alerte « clients »
  - Website Watcher <http://aignes.com>
  - Kbcrawl <http://www.beaconseil.com>
- Certains « aspirateurs de sites » font office d'agents d'alerte
  - Wysiogot <http://www.wysigot.com>

19e journée RNDH - 21/09/2006

## Les agents d'alerte : fonctions avancées

- Limitation des alertes non pertinentes
- Critères de modification avancés
  - nombre de phrases modifiées,
  - lien(s) modifié(s),
  - pourcentage de contenu modifié,
  - images,
  - page disparue
- Extraction des modifications de données

19e journée RNDH - 21/09/2006

## Alertes récurrentes

- [Google alertes](#)
  - Service du moteur de recherche Google (bêta)
  - Permet d'effectuer une veille sur 4 critères:
    - Actualités
    - Web
    - Actualités & Web
    - Groupes
  - Envoie un mail à chaque fois qu'il y a une nouveauté
  - Nécessite la création d'un compte et d'une identification
- [Googlealert](#) (indépendant de Google)
- [Yahoo alerts](#)
- etc.

19e journée RNDH - 21/09/2006

## Les desktops

- L'information produite en interne n'est pas à négliger dans un processus de veille
- Les desktops sont des moteurs de recherche qui indexent le contenu de votre ordinateur
  - [Exalead one desktop](#)
  - [Google desktop](#)
  - [Windows Desktop Search](#)
  - [Yahoo desktop](#)
  - etc.
- Pour des applications professionnelles (réseau) => Solutions payantes

19e journée RNDH - 21/09/2006

## Les outils spécialisés

- [ChemFinder](#) : chimie
  - Interroge des bases de données gratuites et payantes
- [Medhunt](#) de Health on the net : permet de trouver des sites Web médicaux
- [Scirus](#) : moteur de recherche scientifique
- [Sciseek](#) : moteur de recherche scientifique
- [Google scholar](#) : articles, conférences, etc.
- [Yahoo! Education](#) : ouvrages de références

19e journée RNDH - 21/09/2006

## Les blogs : un outil pour la veille ?

- « Journal en ligne » sur Internet ou Intranet tenu par une ou plusieurs personnes
- Possibilité d'insérer des liens, d'ajouter ses commentaires, ses points de vue, ses activités...
- Suivi de l'évolution d'une idée, d'un thème de projet
- Estimation : 40 à 60 millions de blogs dans le monde

19e journée RNDH - 21/09/2006

## Intérêt des blogs

- Collecte d'information sur des **sujets émergents** : R&D...
- **Identification d'experts**, de passionnés d'un sujet
- Exploitation des [commentaires](#)
- **Trackbacks** (permet de relier des articles sur le même sujet postés dans des blogs différents) et **permalien** (url permanente d'un billet sur le site)
- Exploitation des **blogrolls**
- Information catégorisée
- Information datée
- Accès aux archives

19e journée RNDH - 21/09/2006

## Outils de recherche de blogs

- Annuaire
  - [1001rss.com](#)
  - [RSS Network](#)
- Moteurs
  - Les moteurs généralistes : [Google](#), [Live.com](#), etc.
  - Les moteurs spécialisés : [Icerocket](#), [Technorati](#), etc.

19e journée RNDH - 21/09/2006

## Syndication de contenu : les fils RSS

- **RSS** : Really Simple Syndication ou Rich Site Summary
- Format d'échange ([XML](#))
- [Mode de diffusion](#)
  - permet de publier des titres de nouvelles ou d'articles, exploitables dynamiquement par d'autres sites ou par des agrégateurs
- Agent d'alerte
  - permet d'être alerté en permanence sur un domaine d'actualité ou sur les nouveautés apparaissant sur un site précis
  - nécessite un lecteur de fils RSS

19e journée RNDH - 21/09/2006

## Les lecteurs de fils RSS

- Lecteurs en ligne
  - [Bloglines](#)
  - [Netvibes](#)
  - etc.
- Lecteurs clients
  - [Feedreader](#)
  - [RSS bandit](#)
  - [Thunderbird](#)
  - etc.

Dans les deux cas, nécessité de s'abonner

19e journée RNDH - 21/09/2006

## Fils RSS

- Actualités
  - [Le Figaro](#)
  - [Le monde](#)
  - Libération
- Administration (concours, etc.)
  - [Service-public.fr](#)
  - etc.
- Europe
  - [Europa](#)
- Législation
  - [Assemblée nationale](#)
  - [Senat](#)
- Veille juridique
  - [Droit Zoom](#)

19e journée RNDH - 21/09/2006

## Fils RSS

- Brevets
  - [Fresh Patents](#) (catégorie médecine)
- Revues scientifiques (anglais)
  - [Electronic Journals](#)
- Médecine
  - [Environmental Health Perspectives](#)
  - [PubMed](#) => [hubmed](#)

19e journée RNDH - 21/09/2006

## Avantages et inconvénients des fils RSS

- **Avantages**
  - **Gain de temps**
  - **Segmentation** : possibilité de s'abonner à une rubrique spécifique d'un journal
  - **Anonymat** : contrairement à une liste de diffusion, il n'est pas nécessaire de s'identifier
  - **Consultation aisée** : identique aux mails
- **Inconvénients**
  - **Perte de temps** si on ne sélectionne pas correctement nos sources
  - **Redondance de l'information** : cherchez la source
  - **Validité de l'information** : publication plus aisée, n'importe qui peut publier n'importe quoi

19e journée RNDH - 21/09/2006

## Et demain ?

- L'avenir des agents **d'alerte** : leur intégration dans des solutions logicielles
- Les choix : **agent logiciel ou plate-forme Web** ?
- L'évolution des **normes et standards**
- L'**accès aux contenus à valeur ajoutée** : gérer la redondance et les coûts d'abonnement

19e journée RNDH - 21/09/2006

## Contact

Tél. : 06 80 74 41 62  
j.boustany@laposte.net  
<http://www.docinfos.com>  
<http://search.docinfos.com>

19e journée RNDH - 21/09/2006