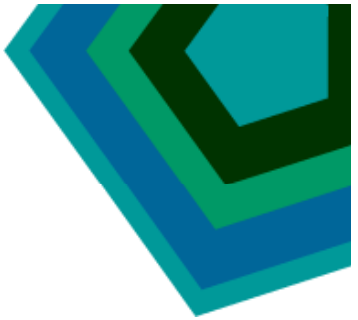
Frédéric Martinet

Veille et Recherche d'informations



Veille et recherche d'informations

2005 - 2006

Définitions

- Veille
  - **AFNOR :** « Activité continue et en grande partie itérative visant à une surveillance active de l'environnement technologique, commercial, etc., pour en anticiper les évolutions » Source: Association française de normalisation, 1998)

Définitions

- Veille informationnelle
  - « On peut définir la veille informationnelle comme **l'utilisation de moyens technologiques** pour **connaître les éléments et les mouvements stratégiques et opérationnels de l'environnement des organisations**. Conséquemment, la veille informationnelle s'adapte à la nature de l'entreprise par un **cadre d'organisation formelle**. Le déluge d'informations maintenant disponibles par le biais des inforoutes doit être endigué dans un moule soigneusement défini au préalable. L'activité de veille est complexe et demande une extrême rigueur mais aussi une intuition particulière souvent issue d'une large connaissance de la culture de l'organisation et de son secteur d'activité. Une veille efficacement structurée permet de prédire avec précision le temps qu'il fera dans un secteur d'activité. Elle a pour objectif de donner une **information ponctuelle; pertinente; vérifiée et synthétisée aux décideurs stratégiques de l'organisation.** » (Source: Monique Fréchette, http://www.itinerant.qc.ca/syndicalisme06.html)

La Veille pour quoi faire?

- Choisir un positionnement concurrentiel
- Détecter des pistes d'innovation
- Faciliter l'accès à l'information en éliminant le superflu et en améliorant le circuit de diffusion
- Alimenter les décideurs en information pour éclaircir leur perception de l'environnement et faciliter la prise de décision stratégique
- Mieux connaître son environnement et en anticiper les tendances
- Identifier de nouvelles pistes de développement commercial

Importance de la veille

- Croissance exponentielle de la masse d'information blanche
- Internationalisation de l'activité économique
- Environnement économique très fluctuant
- Evolutions rapide des normes, lois, application du principe de précaution

Définitions

- Information Scientifique et Technique
  - Éléments de connaissance susceptibles d'être représentés à l'aide de conventions pour être conservés, traités ou communiqués [éléments de connaissance émanant uniquement de l'activité scientifique et technique]
  - Élargissement à l'Information Scientifique, Technique et Professionnelle au XXème siècle : éléments économiques, financiers et commerciaux

Définitions

- Source primaire
  - Publication à l'origine d'une information (article, brevet…)
- Source secondaire
  - Permet la diffusion d'une information, analyse cette information, compilation et « amélioration » par la mise en perspective

Plusieurs typologies d'information

- Information informelle / formelle
- Information stratégique / tactique / opérationnelle
- Information blanche / grise / noire

Formelle / Informelle

- Information formelle
  - Information écrite
  - Informations qui ont un support : papier numérique
  - Journaux spécialisés, revues professionnelles, magazines, cours, articles scientifiques, brevets, bases de données scientifiques et documentaires, Internet
  - Plutôt veille
- Information informelle
  - Information sans support
  - « Ce qui est dit »
  - Partenaires de l'entreprise, Réseaux d'experts internes/externes, Reverse-engineering, Foires et salons
  - Plutôt IE

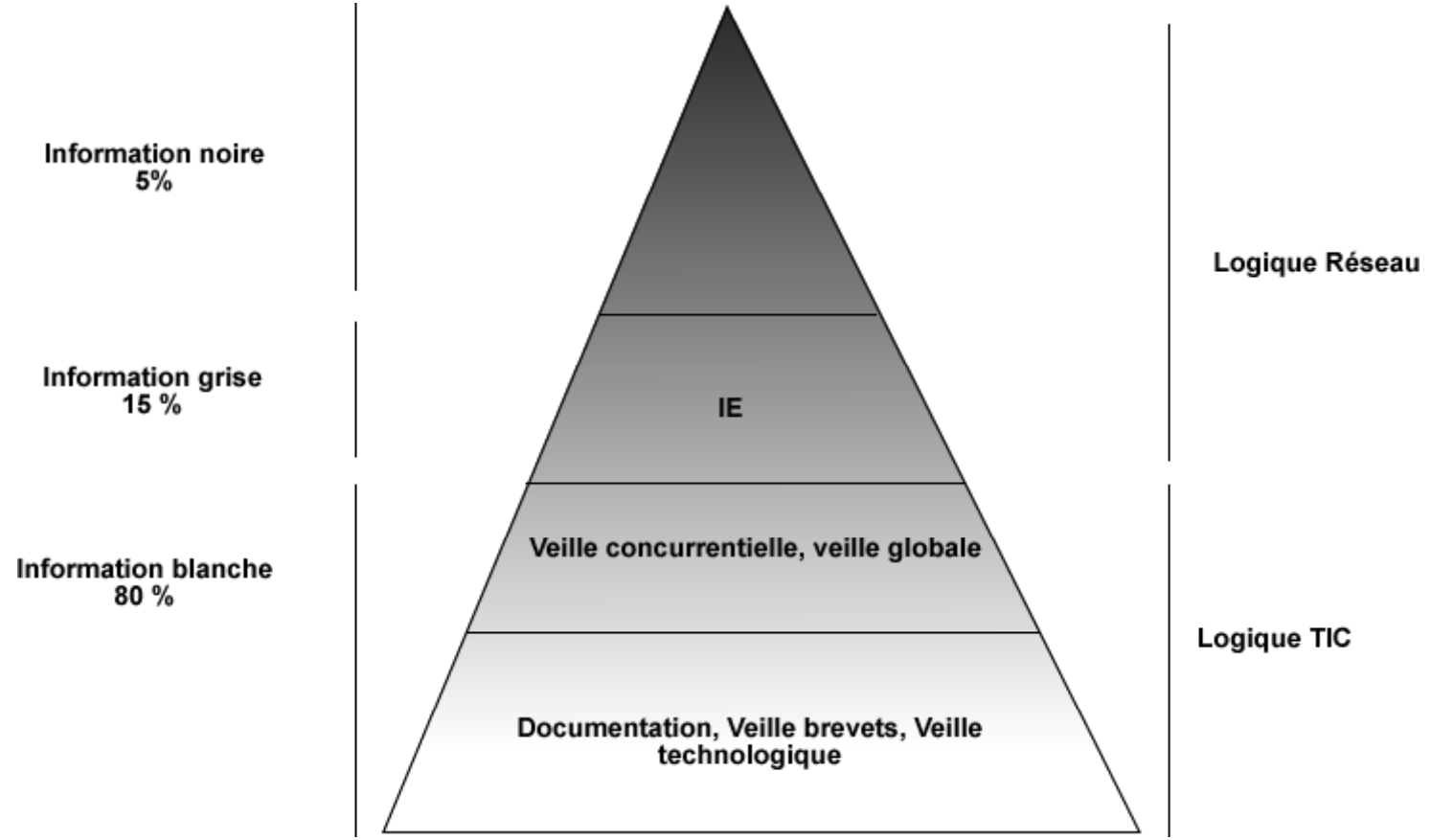
Opérationnelle / Tactique / Stratégique

- Information opérationnelle
  - Information très ciblée, précise, de faible volume
  - Information brute ou peu retraitée
  - A destination des techniciens, ingénieurs, opérateurs
  - Veille brevet et technique par exemple
- Information tactique
  - Volume d'informations plus élevé
  - Subissant un lourd traitement
  - Veille concurrentielle, tarifaire
- Information stratégique
  - Ne concerne pas directement l'entreprise
  - De grande ampleur
  - Permet d'établir des indicateurs
  - Veille marché, prospective, statistiques
    - Cf Mémoire Valérie Léveillé - Université Aix Marseille – 17/01/2000

Information blanche / grise / noire

- Information blanche
  - Facilement accessible par tous
  - Peu de valeur
  - Nécessite tri et traitement important
  - Veille, bibliométrie, data-mining
- Information grise
  - Information difficilement accessible
  - A forte valeur
  - Souvent informelle
  - Indiscrétions, salons, …
  - Intelligence économique
- Information noire
  - Information ne pouvant être acquise que de façon illégale
  - Information décisive pour l'entreprise
  - Espionnage industriel

Logique d'acquisition et Information



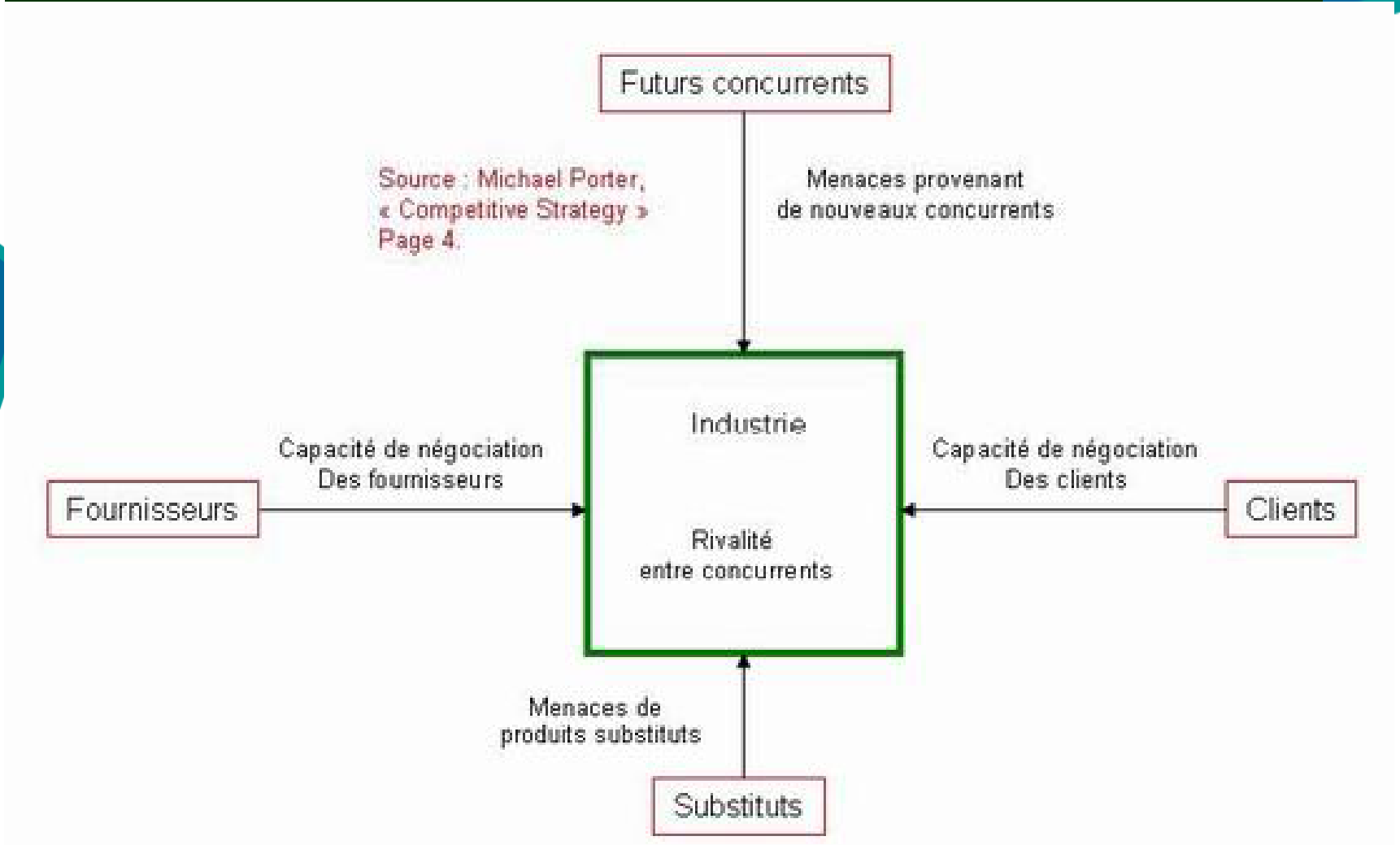
L'intelligence économique, 2ème édition, Alain BLOCH Economica

Valeur de l'information

« La valeur de l’information est égale à la différence entre le bénéfice attendu d’une décision prise sans l’information et celui attendu d’une décision prise avec cette même information. »

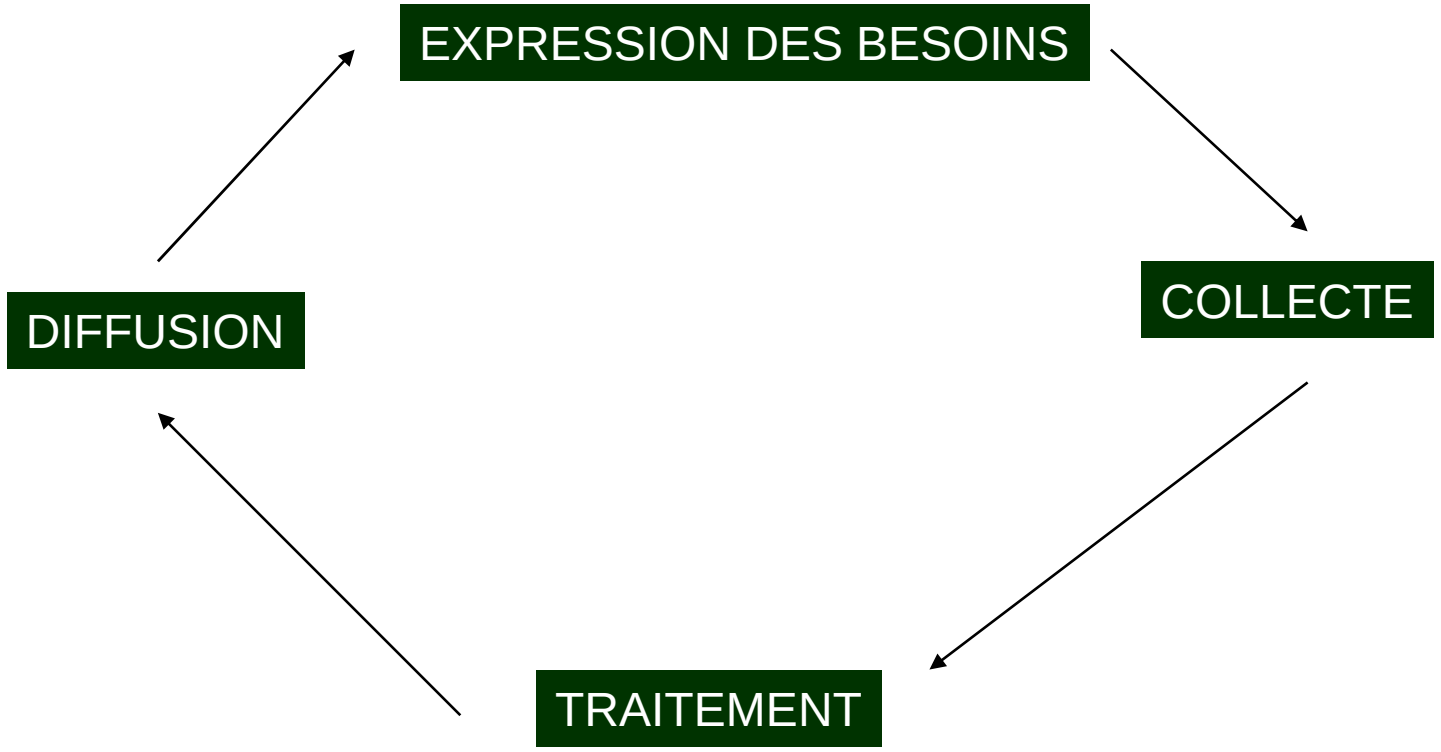
(Source: Huber, 1980, cité dans Taylor, *Value-added Processes in Information Systems*, 1986)

Schéma de Porter et veille



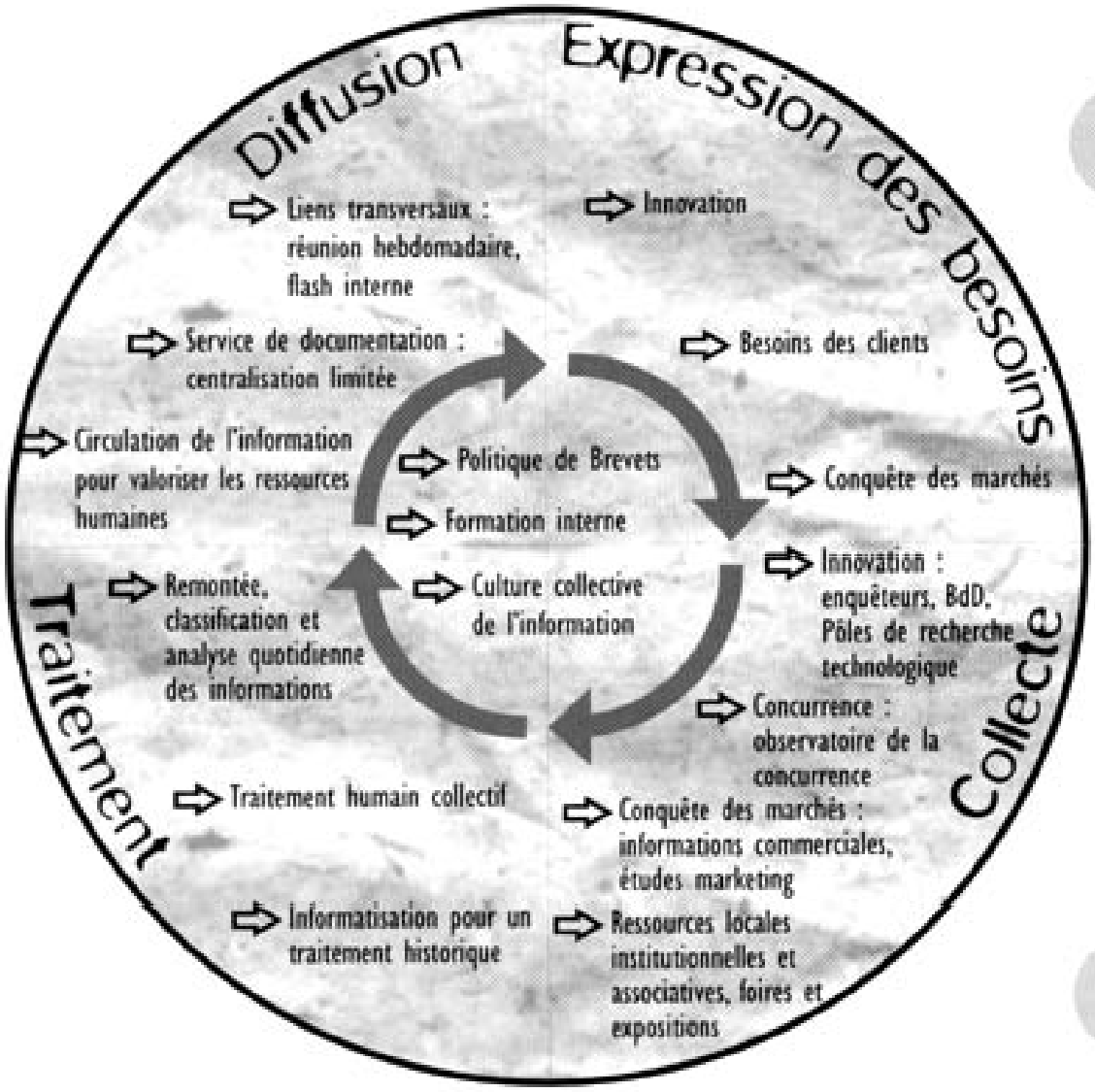
Voir aussi : http://www.ext.upmc.fr/urfist/archives/aurelie/Tableau/Tporter.htm

Le cycle du renseignement



http://www.guerreco.com

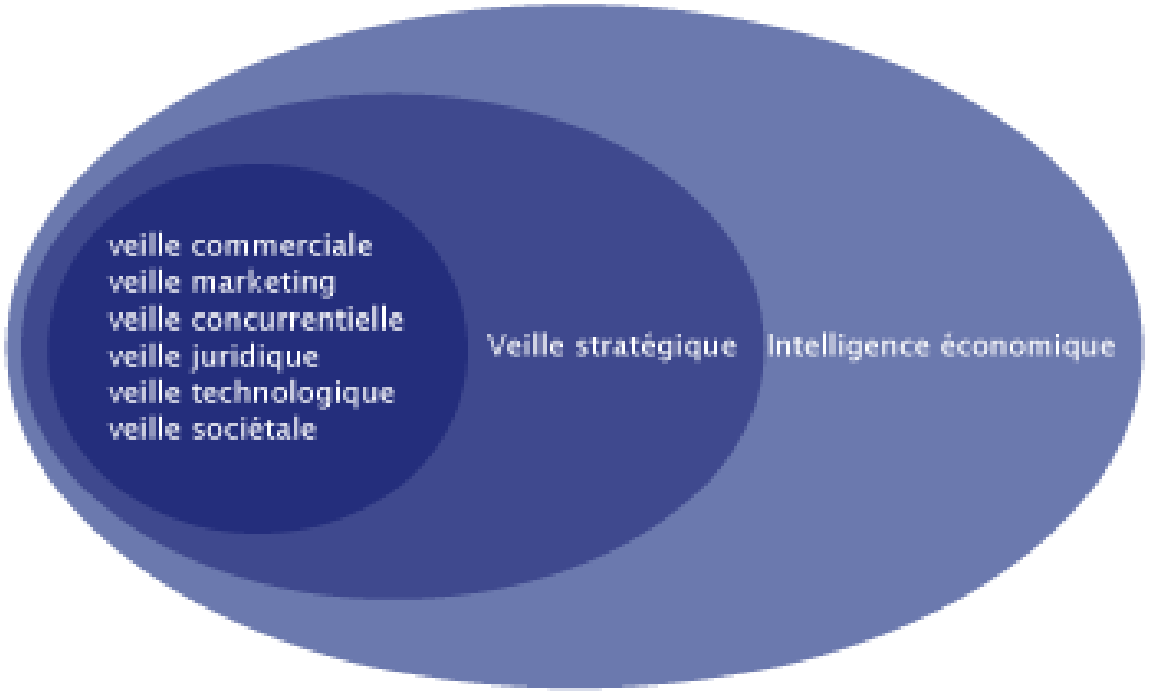
Les PME face au défi de l'Intelligence Economique, Laurent Hassid, Pascal Jacques-Gustave, Nicolas Moinet



Le cycle du renseignement : points clés

- Notion de récursivité
- Il s'agit d'un cercle vertueux
  - Besoins initiaux alimentent et structurent le processus d' IE
  - L'information collectée et diffusée fait évoluer les besoins
  - Emergence de nouveaux besoins
- Importance du feedback !
  - Mettre en place des éléments d'appréciation par l'utilisateur
  - En tenir compte

Les veilles et l'Intelligence économique



Source : http://www.doubleveille.net/intelligence_economique.htm

Intelligence Economique – Rapport Martre

- Différentes définitions qui ont évolué avec les époques et avec l'environnement économique et les relations internationales
- Formalisation de l'Intelligence économique : Rapport Martre - Commissariat du Plan, "Intelligence économique et stratégie des entreprises" (La Documentation Française, Paris, 1994)
  - L'intelligence économique peut être définie comme l'ensemble des actions coordonnées de recherche, de traitement et de distribution, en vue de son exploitation, de l'information utile aux acteurs économiques. Ces diverses actions sont menées légalement avec toutes les garanties de protection nécessaires à la préservation du patrimoine de l'entreprise, dans les meilleures conditions de délais et de coûts. L'information utile est celle dont ont besoin les différents niveaux de décision de l'entreprise ou de la collectivité, pour élaborer et mettre en œuvre de façon cohérente la stratégie et les tactiques nécessaires à l'atteinte des objectifs définis par l'entreprise dans le but d'améliorer sa position dans son environnement concurrentiel. Ces actions, au sein de l'entreprise, s'ordonnent autour d'un cycle ininterrompu, générateur d'une vision partagée des objectifs de l'entreprise«
  - Première définition essentielle qui pose les limites éthiques et légales de l'IE
  - Définition opérationnelle qui dessine un premier schéma organisationnel de l'IE : Recherche / Traitement / Diffusion / Utilisation

Définitions - AFDIE

- «L'intelligence économique peut se définir comme la capacité de l'entreprise à combiner efficacement les réseaux et compétences extérieures en vue de résoudre un problème productif inédit » - 1997 – G. Colletis – AFDIE – Revue d'intelligence économique numéro 1
- Met en avant la capacité des acteurs à se coordonner mais aussi à identifier les personnes et les compétences le tout dans une seule logique : produire de façon nouvelle ou un produit/service nouveau.

Intelligence Economique – Rapport Carayon

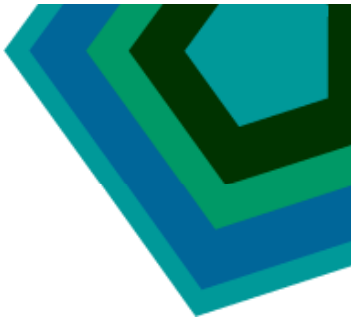
- Vers une nouvelle définition de l'IE
  - « Elle peut nous permettre d'anticiper, l'avenir, de définir ce qu'il est essentiel de promouvoir et de maîtriser notre destin,[…] de définir une politique dans laquelle l'industrie, […] créatrice d'emplois, retrouverait le rang de priorité nationale .»
  - « L'IE devrait être une grande politique publique de l'Etat. »
  - « L'IE est un patriotisme économique.[…]Le patriotisme économique est une politique sociale. »
    - L'IE est ici présenté comme une politique nationale. On voit ici apparaître clairement la défense économique étatique ce qui est inédit.
    - Cf Rapport Carayon – Intelligence économique, compétitivité et cohésion sociale – Documentation Française – Juillet 2003

Définitions – Alain Juillet

- L'intelligence économique consiste en la maîtrise et la protection de l'information stratégique pour tout acteur économique. Elle a pour triple finalité la compétitivité du tissu industriel, la sécurité de l'économie et des entreprises et le renforcement de l'influence de notre pays.
  - Définition qui insiste sur « l'influence nationale » et sur le rayonnement économique
  - Rappel de l'importance de la protection de l'information (cf aussi le rapport du Cigref 2005) , du rôle des DSI, de la sécurité informatique avec une volonté de franciser les outils informatiques

Frédéric Martinet

Veille et Recherche d'informations



Internet pour la veille

Structure du web

Genèse

- Né d'un besoin d'échange entre chercheurs et armée
  - Courrier électronique – 1972
  - FTP – 1973
  - Telnet – 1974
  - Usenet - 1979
  - Internet Relay Chat (IRC) – 1988
  - Peer to peer - 2000

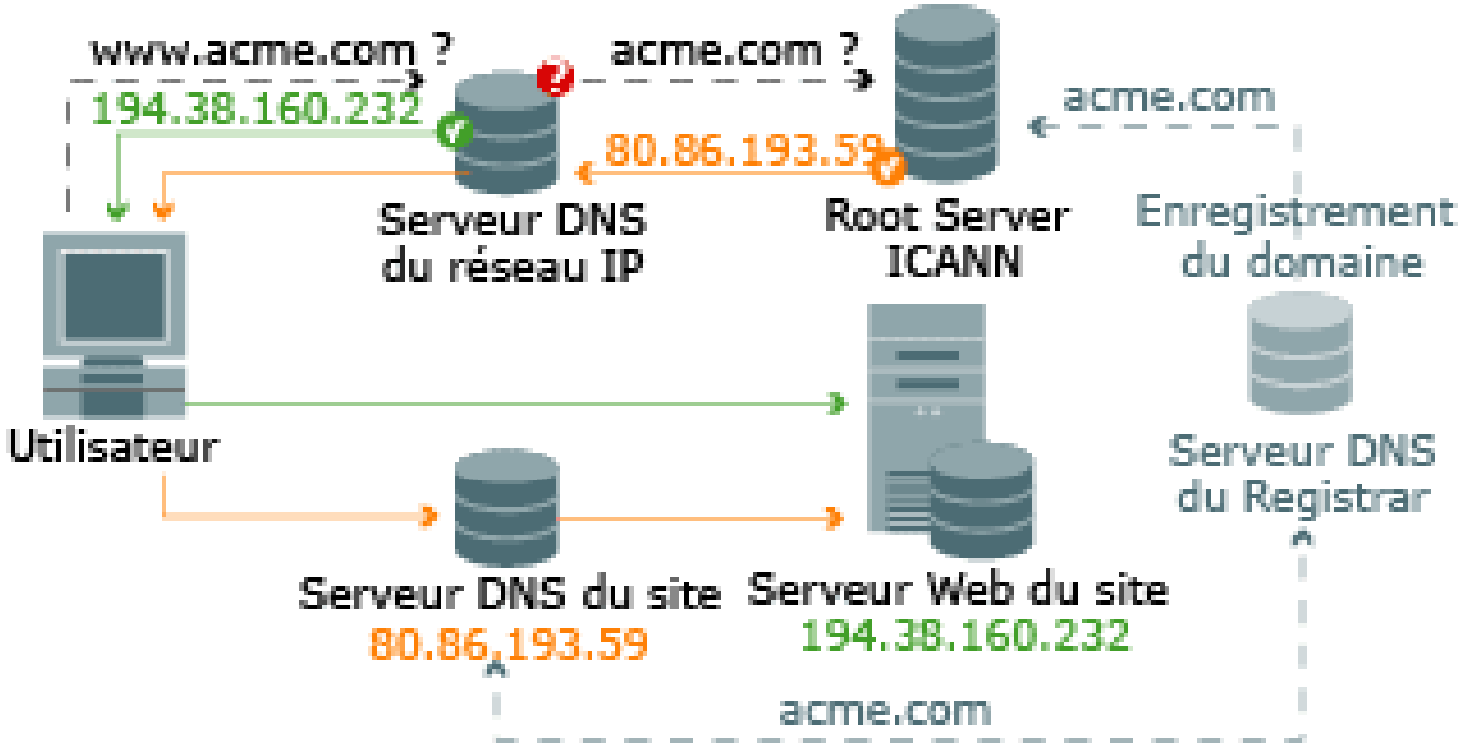
Internet et le contenu

- Au départ, donc, besoin d'échange de données
  - Internet outil de diffusion du contenu?
    - Pendant longtemps
    - Evolution des débits
      - Encore en 1999 le 56 k était la généralité (en France)
      - La démocratisation des débits a peu à peu révolutionné le Web
      - Plus de contenu…plus de services…plus de graphismes…
    - Evolution vers l'utilisation du Réseau à d'autres fins : télé, VOIP, Téléchargements multimédias…

L’architecture du « Net »

- Des milliards de page webs
  - Aucune évaluation de taille possible
  - Web visible et invisible
  - Google indexe 8 168 684 336 pages (septembre 2005) – [dernière évaluation officielle disponible]
- Des sites hébergées sur des machines différentes
  - Hébergement mutualisé
  - Hébergement dédié
  - Sur ordinateurs personnels
  - Comment mettre en lien une machine souhaitant accéder à un site et le site lui-même?
    - L’URL (Uniform Resource Locator)
    - Des DNS (Domains Name Systems)
    - Les adresses IP (Internet Protocol)

Résolution DNS



1. Le serveur DNS du réseau IP de l'utilisateur connaît déjà l'adresse du site web.
2. Le serveur DNS du réseau IP ne connaît pas encore le site et reçoit d'abord d'un Root Server l'adresse du serveur DNS qui gère le site web.

Résolution DNS

- 13 serveurs racines dans le monde
  - http://solutions.journaldunet.com/0304/030416_faqdns.shtml
  - http://en.wikipedia.org/wiki/Domain_Name_System
- En pratique ne répondent qu'à une infime partie des requêtes
  - Grâce à des serveurs tampons dupliqués (Domain name resolvers)
  - Grâce au cache des serveurs locaux

Insertion des langages

- Avant pages Internet « figées »
  - Le contenu et la forme sont liées
  - Toute modification du contenu passe par des modifications « longues » et nécessitant des logiciels
- Evolution vers un Internet à fort contenu
  - Le navigateur Internet devient le meilleur ami du webmaster ou du « gestionnaire de contenu »
  - Les données sont stockées dans une base
  - Le site appelle les données et construit la page en fonction des requêtes (clics, recherche, interactions diverses)
- On distingue désormais le web statique (HTML) du web dynamique (PHP, ASP, JSP)

La page Web

- Le langage initial du web : le HTML
  - Version actuelle 4.01
    - http://www.w3.org/TR/html4/
  - A conquis par sa simplicité!
  - Des insuffisances : peu d'interactivité, très statique, simple mise en forme et liens hypertextes entre les documents
- D'autres langages : DHTML, Javascript, XML, CSS….viennent compléter le HTML
- D'autres technologies : Flash

Structure d’une page Web

- <HTML><HEAD></HEAD><BODY></BODY></HTML>
- Entête : Contient les informations d’indexation documentaire : les META
  - Et des éléments de codes : javascripts, css, etc
  - Title
  - Description
  - Keywords
  - Expires
  - Author
  - Reply to

Pourquoi des META?

- Web => Information « brute », non catégorisée, non ordonnée
  - Nécessité de meilleure lisibilité du contenu
  - Historiquement utilisées par les moteurs de recherche
  - De plus en plus désuètes sauf la title
    - A cause du spamdexing : détournement de ces balises pour frauder les moteurs
    - Grâce à l'accroissement de la puissance de calcul
    - Grâce aux nouveaux algorithmes de catégorisation automatique

Le web statique

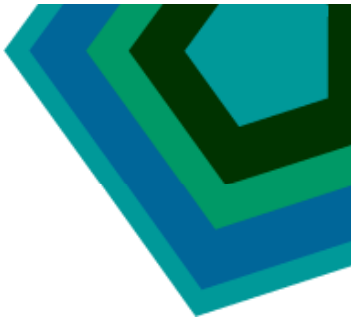
- En HTML
  - Contenu et forme imbriqués
  - Toute modification du site nécessite
    - Le passage par un éditeur WYSIWYG
    - La mise à jour du site via FTP
  - Procédure lourde
  - Procédure technique
  - Entraînait des contenus peu mis à jour, une information plus souvent périmée

Le web dynamique

- PHP, ASP…
  - Contenu et forme dissociés
  - Forme : HTML + CSS, Contenu : SQL + PHP
  - Modification du ligne via un simple navigateur (pour le rajout de contenu)
  - Procédure simple
  - Immédiateté de l'information
    - Avantages et inconvénients : l'information se multiplie, disparaît facilement

Frédéric Martinet

Veille et Recherche d'informations



Evaluer l'information

Vérifier l'information

- Estimation des sources
  - Fiabilité du média
  - Fiabilité de l'auteur et spécialisation
- Croisement
  - 3 canaux de diffusion différents
  - Difficulté sur Internet et pour l'information numérique
  - Attention aux ressemblances, fautes d'orthographe identiques dans deux mêmes infos
  - Attention aux articles trop orientés commercialement
  - Intoxication par dépôt de brevet possible
    - Grille d'évaluation d'un site WEB
    - Quelques questions à se poser

Critères traditionnels

- Exactitude
  - Fiable / exempte d'erreurs /contrôlée
- Autorité intellectuelle
  - Qualification / réputation de l'auteur
- Objectivité
  - Minimum de préjugés / tente d'influencer le lecteur
- Actualité
  - Mise à jour / Date de publication
- Couverture
  - Sujets ? / Traités en profondeur
    - Source : www.media-awareness.ca

Problématiques Internet

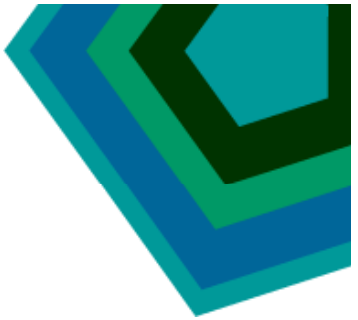
- Sites webs commerciaux
- Infopubs
- Liens hypertexte
- Limitation d'accès à l'information
- Pages webs hors contexte
- Altération des pages web
  - A voir : http://www.widener.edu/Tools_Resources/Libraries/Wolfgram_Memorial_Library/Evaluate_Web_Pages/659 pour les anglophones
  - http://sosig.ac.uk/desire/internet-detective.html

Exercices

- Evaluer exactitude et autorité intellectuelle
  - http://www.copains-canins.com/index.php3/milieu/voir/id/74/table/infos
  - www.jacqueschirac.org - !
- Evaluer l'objectivité
  - http://www.protegez-vous.qc.ca/
  - http://www.dhmo.org

Frédéric Martinet

Veille et Recherche d'informations



Les moteurs de recherche

Indexation par les spiders

- Crawl du web
- Parsing =>indexation du document ou mise à jour de l'index
- Une partie du document
  - Balises, début du contenu, titre, URL
  - Suppression des mots vides
  - Création d'un gigantesque index inversé
    - Gain de temps

Crawl

- Crawl : surf par les moteurs de recherche
  - Les robots ou spiders sont des agents intelligents développés par les moteurs de recherche
  - Ils vont de page en page en suivant les liens hypertexte
  - Ils sont régis par :
    - Les règles du programmeur
    - Les règles Internet
  - Ils indexent les pages
  - http://outils.abondance.com/

Quelques règles

- Profondeur de crawl
- Type de documents
- Se conforme à robots.txt
  - http://www.robotstxt.org/wc/exclusion.html#robotstxt
  - <META NAME="robots" CONTENT="index,nofollow">
- Fréquence de crawl
- Balises indexées
- Poids maximum du document ou limite de poids indexé
- …
- Beaucoup de règles inconnues : le fonctionnement des spiders sont inconnus afin de ne pas être contournés
- Nom des robots :
  - http://www.searchengineworld.com/spiders/spider_ips.htm
  - http://www.robotstxt.org/wc/active/html/index.html
  - http://www.mjbdata.co.uk/spiders/
  - http://joseluis.pellicer.org/ua/

Les requêtes

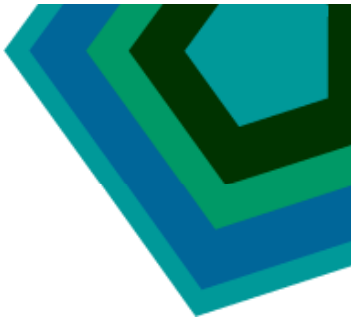
- Interrogation des moteurs de recherche par une interface de type formulaire
- Disparition des opérateurs booléens et du langage d'interrogation type « base de données professionnelles »
- Souvent interface de type recherche avancée avec options complémentaires
- La précision de la requête est remplacée par le calcul de la pertinence

Pertinence

- Boîte noire propre à chaque moteur
  - Certaines informations sont disponibles d'autres non afin de préserver du spamdexing
  - La plupart du temps
    - Recherche des documents avec tous les mots
    - Importance de la proximité des mots
    - Importance de l'ordre
    - Importance du nombre d'occurrences

Frédéric Martinet

Veille et Recherche d'informations



Focus Google

Google - Présentation

- Moteur de recherche généraliste
- Marque internet bénéficiant de la plus grande notoriété
- 8 milliards de page indexées, index en augmentation
- Garde une trace en cache des informations indexées
- Documents html, doc, pdf, xls, mdb, dwg, ps, ppt, rtf, xml

Le concept

- A démocratisé la notion de page rank
  - Chaque PAGE bénéficie d'une popularité
    - Qui dépend du nombre de liens pointant vers cette page
    - Qui dépend de la popularité des pages pointant vers cette page
  - Se calcule de façon récursive
    - PR(B) = (1-d) + d x (PR(A1) / N(A1) + ... + PR(An) / N(An))
    - http://www-db.stanford.edu/~backrub/google.html ou plus simple http://www.webmaster-hub.com/publication/article16.html et son évolution Documents\DeeperInsidePR.pdf

Google - Le calcul de pertinence

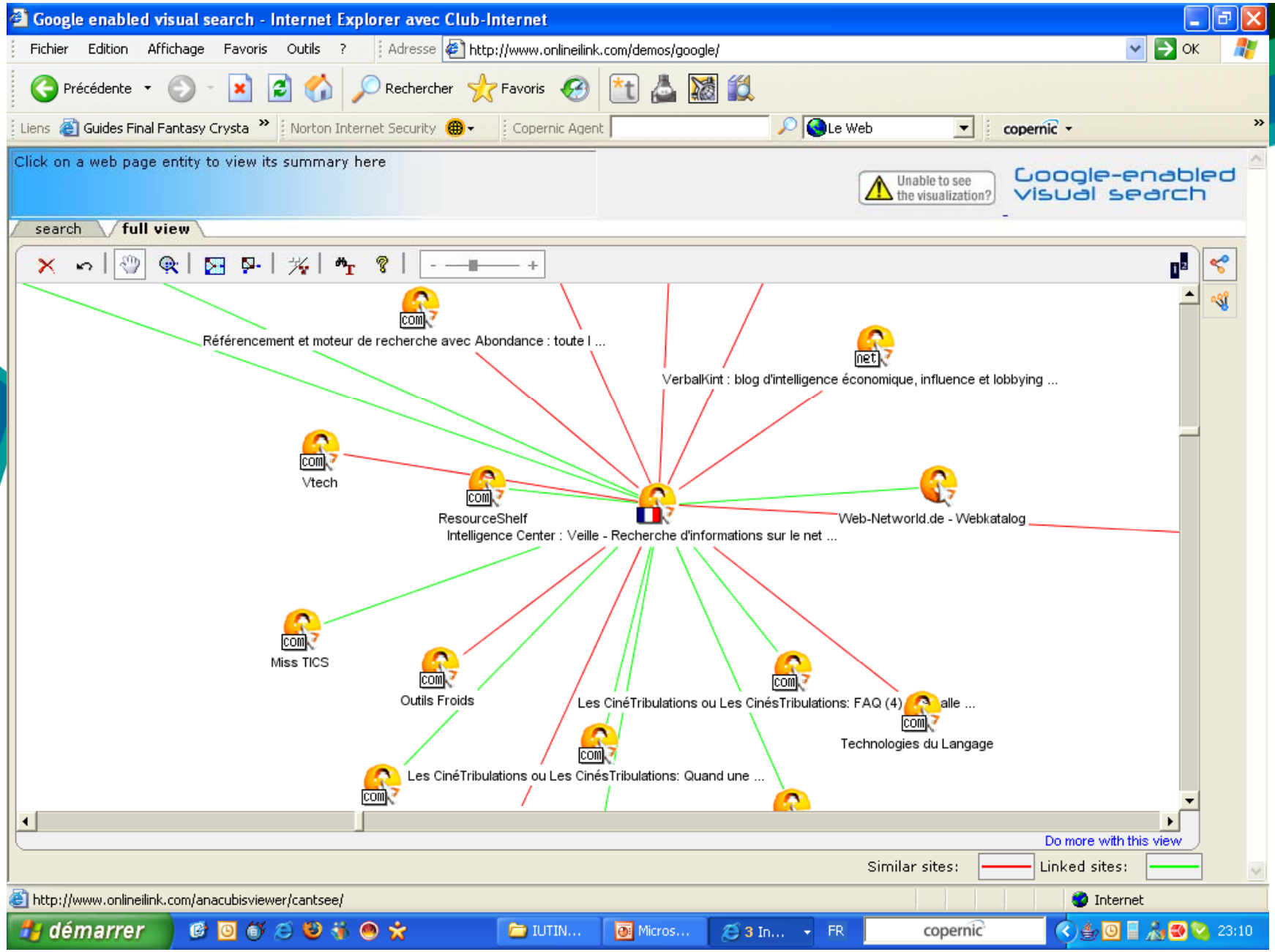
- Envoi de mot(s) clé(s) à Google
- Sélection de toutes les pages liés à ce mot dans l'index inverse
- Calcul de la pertinence pour chacune des pages
  - Prise en compte du nombre d'occurrences, de la position du mot, **du page rank**, de la pertinence de pages pointant vers cette page sur ce mot clé
  - Cet algorithme est plus ou moins secret
  - Tests permettent d'en savoir plus, de détecter les changements
- Tri des résultats

Les moteurs qui comptent

- Yahoo
- MSN
- All the Web
- Exalead
  - Tous n'intègrent pas la notion de page rank
  - « Simple » calcul de pertinence
  - All the web propose d 'effectuer des requêtes booléennes
- Quaero???!!!!

Quelques moteurs innovants

- Aol
  - Avec clusterisation des résultats
  - Permet d'identifier des sous-domaines de recherche ou des domaines liés
- Technologie Anacubis (cf exemple) carto Google
- A voir aussi : Wisenut, …



Exemples (1)

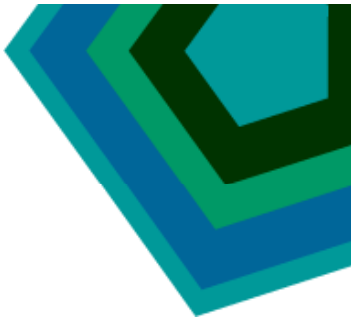
- Requête sur google permettant de trouver les pdf mentionnant « intelligence économique » dans le titre dans des sites en .fr
  - http://www.google.fr/search?as_q=&num=10&hl=fr&btnG=Recherche+Google&as_epq=intelligence+%C3%A9conomique&as_oq=&as_eq=&lr=&as_ft=i&as_filetype=pdf&as_qdr=all&as_occt=title&as_dt=i&as_sitesearch=.fr
- Trouver des fichiers XLS sur le site d'Airbus
  - http://www.google.fr/search?as_q=airbus&num=10&hl=fr&btnG=Recherche+Google&as_epq=&as_oq=&as_eq=&lr=&as_ft=i&as_filetype=xls&as_qdr=all&as_occt=any&as_dt=i&as_sitesearch=airbus.com

Exemples (2)

- Trouver des fichiers powerpoint sur l'université des sciences sociales de Toulouse
  - http://www.google.fr/search?q=ppt+site:univ-tlse1.fr+filetype:ppt&hl=fr&lr=&as_qdr=all&start=0&sa=N
- Trouver des fichiers autocad d'avion
  - http://www.google.fr/search?hl=fr&as_qdr=all&q=filetype%3Adwg+airplane&meta=

Frédéric Martinet

Veille et Recherche d'informations



Les annuaires

Logique

- Logique de site : un site par nom de domaine et pas plusieurs pages
- Classification des sites
  - Par zone géographique
  - Par activité
  - Voir exemple : www.dmoz.fr
- Validation du site avant intégration
  - Par des responsables / éditeurs de rubriques
  - Vérification de l'adéquation entre la catégorie et le contenu

Recherche dans un annuaire

- Une recherche par surf à l'intérieur des catégories
- Recherche par mots clés souvent peu pertinente / trop partielle
  - Une seule page indexée = déclaration faite par les webmasters
- Nécessite une bonne connaissance de l'annuaire et de ses rubriques

Avantages / inconvénients

- Cohérence entre rubrique et site
- Tri sélectif : qualité préférée à quantité
- Impossible de rechercher sur l'intégralité d'un site : restriction à la première page
- Tri sélectif…est aussi un inconvénient : corpus parcellaire, incomplet (plus encore que celui des moteurs)
- Recherches dans les rubriques parfois fastidieuses et difficiles

Dmoz.org

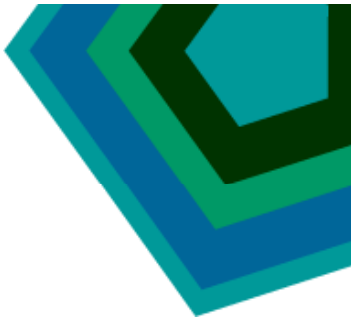
- 4 millions de sites anglais
- 100 000 sites français
- Interrogation sur url (u:), titre (t:), descriptif (d:), catégorie (c:)
- Recherche avancée possible
- Projet open source
- Repris par de nombreux autres annuaires partiellement ou totalement

Yahoo

- Annuaire historique
- fr.dir.yahoo.com/
- Taille d'index non précisée
- Un des premiers annuaires
- Relativement délaissé souffrant de désaffection de la part des internautes
- Développement d'une technologie de moteurs de recherche propre (avant : Google)

Frédéric Martinet

Veille et Recherche d'informations



Les métamoteurs

Fonctionnement

- Recours à plusieurs bases d'index et plusieurs technologies de classement, de pertinence
- Envoie la requête vers les différents moteurs de recherche
- Injecte les résultats dans son algorithme de pertinence
  - Poids variable des différents moteurs
  - Positionnement pondéré de la page
  - Nombre d'occurrences
- Dédoublonne les résultats
- La plupart du temps pas de base propre

Offline / online

- Métamoteur en ligne sur un site web
  - Obligation d'être connecté même après une requête
  - Difficulté de stockage des résultats
  - Information toujours actualisée
  - Souvent gratuit
- Métamoteur offline installé sur le poste client
  - Possibilité de stockage et d'archivage des différents résultats, d'exportation vers d'autres applications
  - Avoir les droits d'installation
  - Souvent payant

Métamoteurs online

- Ixquick
- Kartoo
- Metacrawler
- Mamma
- Mapstan
- Vivisimo / Clusty

Copernic

- Un des leaders des métamoteurs (offline)
  - Version gratuite et payante
  - Inclus de nombreuses autres fonctionnalités
  - Pertinent et efficace
  - Nombreux paramétrages de moteurs
  - Moteurs de news
  - Reste assez anglophone
  - Google absent de l'index sauf si…

Démonstration

- Les moteurs présents
- Paramétrer une requête
- Paramètre des résultats : nombre
- Trier les résultats
- Surveiller une recherche
- Trouver dans les résultats

Métamoteur

- Doit proposer un algorithme de pertinence efficace, différenciateur
- Doit innover sur la présentation ou sur le concept
  - Cartographie, Evaluation (fooxx.com), Search Tuna
- Doit proposer un nombre de moteurs de recherche « assez important »
- Propose des options limitées de recherche avancée
  - Choix des moteurs requêtés
  - Nombre de résultats par moteur
  - Tous les mots, l'expression exacte

Les différents critères

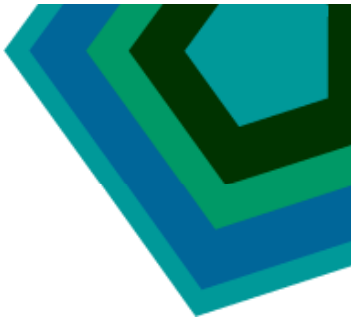
- Nombre de moteurs
- Type de moteurs (générique, news, images brevets…)
- Possibilité de créer ses catégories
- Possibilité de rajouter ses propres moteurs
- Planification de mise à jour des requêtes
- Alerte push
- Surveillance de pages web

Fonctionnement moteurs

Source : lemonde.fr 22/01/2006

Frédéric Martinet

Veille et Recherche d'informations



Quelles sources pour quelle utilisation ?

Surveiller son environnement

- Presse écrite ou en ligne
  - Grands quotidiens
  - PQR
  - Hebdos
  - Médias télévisés
  - Bases de données centralisatrices
- Législation
  - Codes et jurisprudence
    - Facilitation par la mise en ligne
      - » Legifrance.gouv.fr

Surveiller ses concurrents

- Presse…bien sur
- Site web
- Documentation d'entreprise sur les salons
  - http://www.veille.com/fr/article.php3?id_article=27433
- Bases de données d'entreprises
- Information grise fournisseur par exemple

Surveiller les technologies

- Moteur de recherche sur les brevets
  - Français
  - Internationaux
  - Exemples
    - CIPO
    - Esp@cenet
    - UPSTO
    - Plutarque
- Presse professionnelle / spécialisée
- Salons professionnels
  - Attention à l'intoxication

Les outils

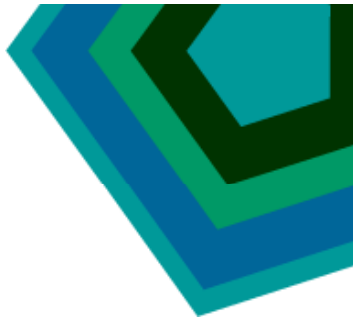
- Informations sur une entreprise
  - Gratuits : societe.com, europages, indexa
  - Payant : ORT, Kompass, DnB, Telefirm, SCRL, Infogreffe, Euridile
- Indicateurs économiques
  - Sites publics : Ministères, Insee, Dree…
  - Ipsos, Xerfi, Dafsa…
- Textes officiels et réglementaires, normes, brevets
  - Journal-officiel.gouv, legifrance, europa.eu.int, inpi, afnor
- Articles, communiqués de presse
  - AFP, sites de presse, europresse, indexpresse

Les serveurs professionnels

- Dialog
  - 15 terabytes, bases internationales
- Questel-Orbit
  - BDD sur les brevets et marques, bases de données scientifiques et techniques
- ORT
  - Information sur les entreprises, Presse, Juridique
- Qwam
  - Portail Dialog, Kompass, Delphes, Xerfi
    - Un exemple de la syntaxe d'interrogation
    - Exemple de doc Dialog

Frédéric Martinet

Veille et Recherche d'informations



La veille

Sourcing

- Préalable : identifier ses besoins, son périmètre
- Tout support
  - Papier
    - Ne jamais négliger les revues professionnelles
    - Les lettres d'information confidentielles
  - Sites webs
    - Listes de discussions professionnelles et / ou spécialisées
    - Newsletters
    - Sites de défense du consommateur
    - Sites associatifs
    - Blogs

Commencer sourcing

- Utilisation des moteurs de recherche et annuaires
  - Thématique : secteur, nom du client, activité
  - Utilisation de la commande link:
  - Utilisation de la commande related:
  - Recherche sur les groupes http://fr.groups.yahoo.com
  - Utilisation des moteurs avec clustering
  - Création d'un bookmark par thématique ou par typologie de site
  - Utilisation de logiciels de gestion des bookmarks
    - http://www.thebrain.com
    - http://www.anshare.com/type.asp?T=53
    - http://inforizon.blogs.com/veille/2004/07/gestionnaires_d.html
    - Mes Favoris – Logiciel de gestion des bookmarks

Comment surveiller ?

- Push
  - Le veilleur reçoit l'information
    - Newsletter après inscription
    - Services de veille spécialisés en ligne
      - Ex : Net2One
    - Paramétrage d'agents intelligents
    - Plateforme de veille spécialisées type Digimind, Knowings
- Pull
  - Le veilleur va chercher l'info
    - Visite régulière de site selon planning
      - 2 logiques : Site peu importants ou site très importants avec visite quotidienne
    - Recherche ponctuelles pour élargir un sourcing initial ou sur une thématique nouvelle

Fréquence

- 2 critères déterminants
  - Fréquence de mise à jour
    - Problème
      - Difficilement connaissable pour certaines sources ou sites
      - Peut varier subitement
    - Connue pour certaines sources
      - Quotidiens! Hebdomadaires! …
  - Fréquence de diffusion de l'information et document de diffusion
    - Alerte : rapidité, nécessité de veille permanente, réactive
    - Rapport à fréquence fixe : plan de veille en correspondance : balayer les sites sur la période

Outils (1)

- Actuellement peut faire correspondre à un besoin un outil
  - Problème du budget
    - Cohérence dans l'importance de veille et son coût
      - ESE industrielle innovante : investissement justifié sur la veille brevets
    - Étude de faisabilité préalable pour mise en place cellule de veille
  - Certains problèmes de bruit sur des thématiques vastes et des problématiques de veille complexes internationales

Outils (2)

- S’inscrivent plus dans une phase sourcing / collecte que traitement / diffusion
  - Diffusion : Blogs / Wikis et autres…
  - Collecte + Diffusion : Plateformes de veilles…
  - Traitement : Text-mining / Data-mining…
- Parfois peu ergonomiques
- Faisant appel à des compétences techniques
  - Voir listing CIGREF 2006 des outils de collecte et de traitement de l’information et rapport 2002 de Fuld

Logiciels de recherche

- Le plus connu
  - Copernic Agent (pro)
- Les défunts
  - Strategic Finder (plus de MAJ) - Digimind
  - BullsEye – Intelliseek (Lire aussi)
- Les challengers
  - Orbiscope Meta Recherche
  - Firstop Websearch

Les Challengers

- Orbiscope
  - Prix faible (40 € approximativement)
  - Possibilité d'ajouter ses propres moteurs de recherche (recherche interne à un site)
  - Peu ergonomique
- Firstop Websearch
  - Catégorisation automatique des résultats
  - Surveillance de page web ou de requête impossible

Agent de veille off line

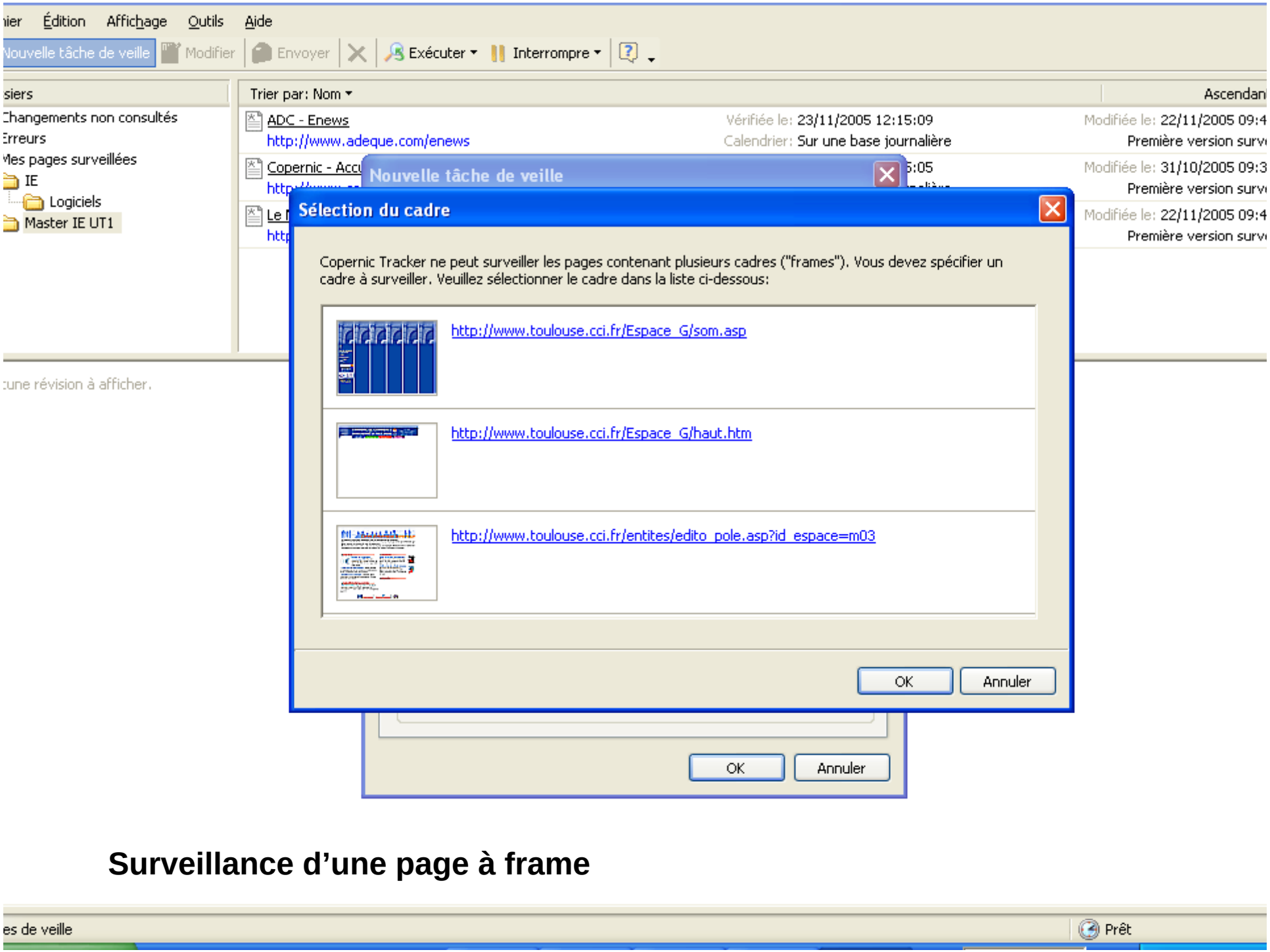
- Installation d'un logiciel sur le poste utilisateur
- Paramétrage des fonctionnalités de push (alerte par mail, smtp etc…)
- Paramétrage pour chacune des requêtes de veille : fréquence, déclencheur de l'alerte
- Alerte par email ou sur le poste par système de pop-up
- Avantages : puissant, archivage des modifications
- Inconvénient : coût à l'achat, pas d'anonymat, nécessité d'installer sur un poste client
  - Website Watcher : http://www.aignes.com
  - Copernic agent : http://www.copernic.com
  - Vigilus smart : http://www.pragtec.com
  - C4U

Copernic Tracker

- Simple d'utilisation
- Surveille les pages uniques : page d'accueil, page spécifique d'un site
- Filtre numériques / dates / mots clés / nombre de mots
- Envoi des alertes par mail à plusieurs destinataires
- Possibilité de surveiller des espaces protégés
- Difficile de diffuser alertes synthétiques : une simple page web surlignée sur les changements envoyée par mail

Exemples

- Paramétrage d'une veille sur une page simple (Copernic)
- Paramétrage d'une veille sur page à forte proportion de contenu régulièrement modifié (Lemonde.fr) : filtre par mots clés
- Paramétrage d'une veille sur une page à frame (http://www.toulouse.cci.fr/Index.asp?fichier=Entites/edito_pole.asp?id_espace=M03)
- Paramétrage d'une veille sur une page protégée par protection serveur (Enews)



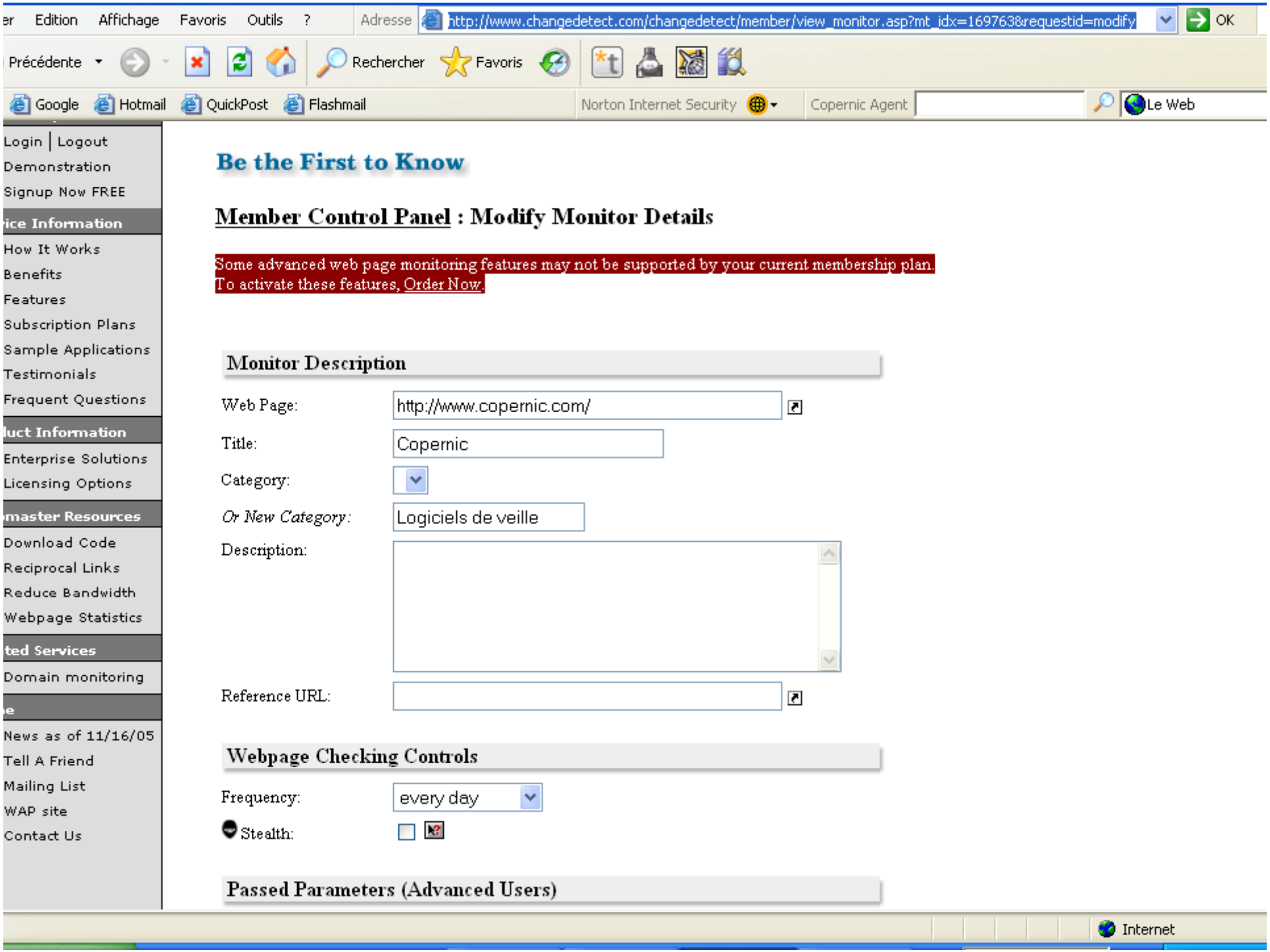
Surveillance d’une page à frame

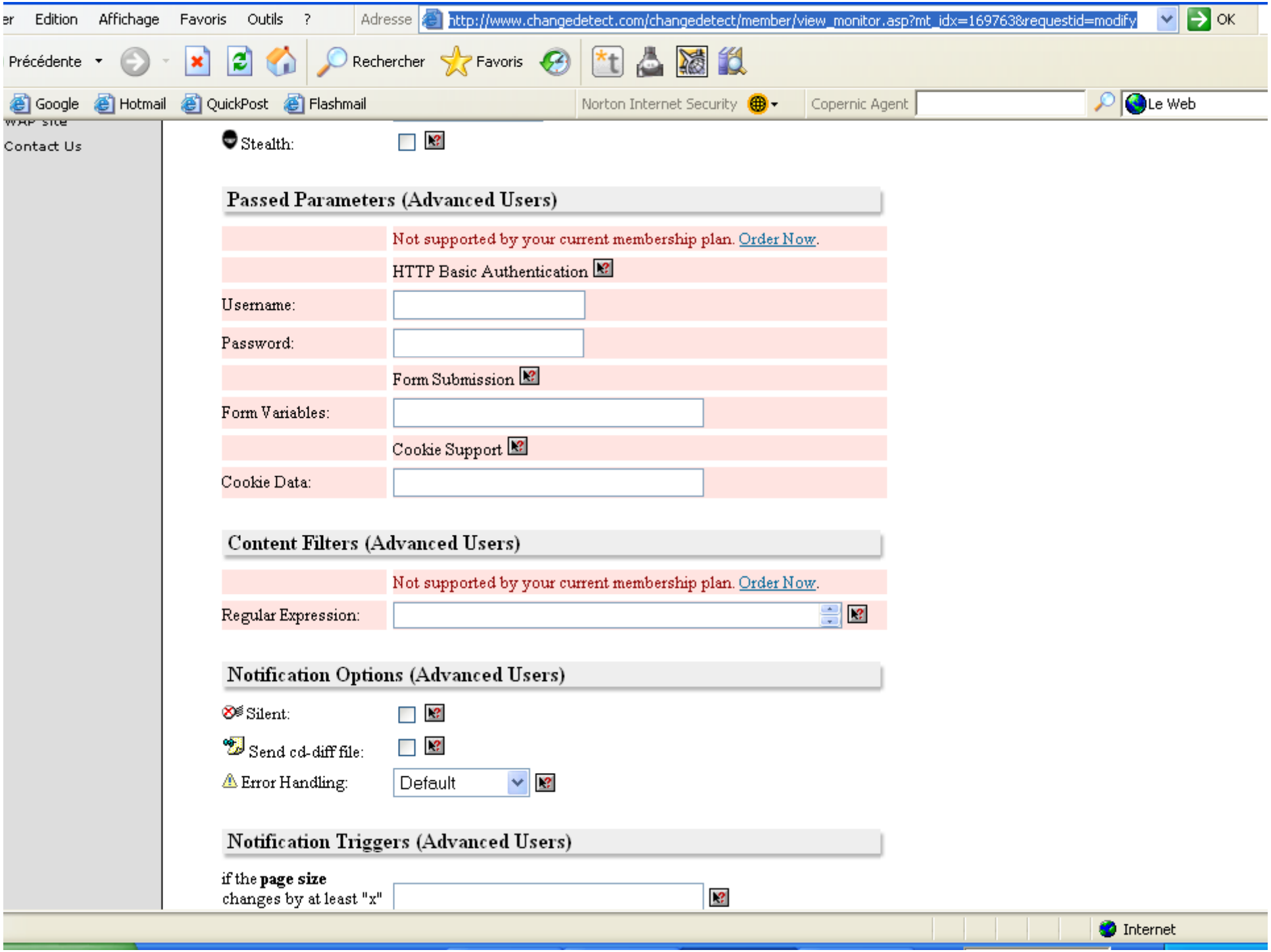
Website Watcher

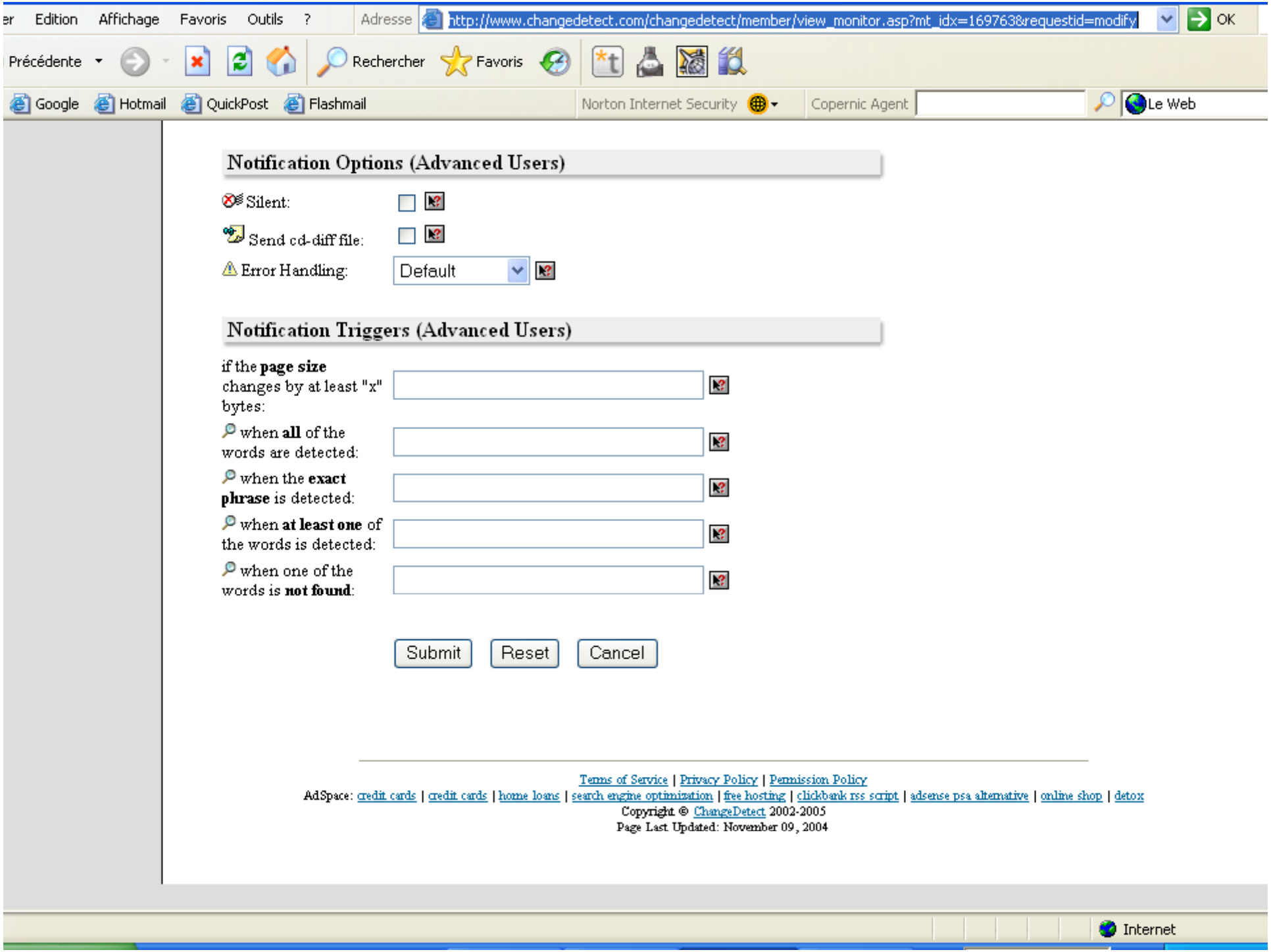
- L'Agent de surveillance page web le plus complet (dans sa gamme de prix)
  - Surveillance de pages et de sites
  - Alertes en push
  - Paramètres de configuration avancés
    - Alerte si mot clé, pas d'alerte si mot clé, planification de veille, gestion de la bande passante, archivage des versions
  - Surveillance de page nécessitant une identification
    - Utilisation des cookies IE
    - Gestion des protocoles Post et Get
    - Gestion des espaces en HTACCESS

Agent de veille en ligne

- Accès à un site web
- Création d'un compte utilisateur
- Eventuellement ajout de barre IE
- Paramétrage des pages et / ou des mots clés
- Alerte par email
- Avantage : anonyme, peu cher, ne nécessite pas d'installation de logiciel
- Inconvénient : Peu pratique, souvent moins efficace que les agents off line, long
  - http://www.snyke.com/ ! ? Change Notes Infominder, Watch that page
  - En disparition….Peu pérennes
  - Plutôt intégré au site par site
- Démonstration Change Detect


er Edition Affichage Favoris Outils ?
Adresse http://www.changedetect.com/changedetect/member/view_monitor.asp?mt_idx=169763&requestid=modify
OK
Précédente
Rechercher
Favoris
Google
Hotmail
QuickPost
Flashmail
Norton Internet Security
Copernic Agent
Le Web
Login | Logout
Demonstration
Signup Now FREE
ice Information
How It Works
Benefits
Features
Subscription Plans
Sample Applications
Testimonials
Frequent Questions
luct Information
Enterprise Solutions
Licensing Options
master Resources
Download Code
Reciprocal Links
Reduce Bandwidth
Webpage Statistics
ted Services
Domain monitoring
e
News as of 11/16/05
Tell A Friend
Mailing List
WAP site
Contact Us
Be the First to Know
Member Control Panel : Modify Monitor Details
Some advanced web page monitoring features may not be supported by your current membership plan.
To activate these features, Order Now.
Monitor Description
Web Page:
http://www.copernic.com/
Title:
Copernic
Category:
Or New Category:
Logiciels de veille
Description:
Reference URL:
Webpage Checking Controls
Frequency:
every day
Stealth:
Passed Parameters (Advanced Users)
Internet


Edition Affichage Favoris Outils ?
Adresse
http://www.changedetect.com/changedetect/member/view_monitor.asp?mt_idx=169763&requestid=modify
OK
Précédente
Rechercher
Favoris
Google
Hotmail
QuickPost
Flashmail
Norton Internet Security
Copernic Agent
Le Web
Contact Us
Stealth:
Passed Parameters (Advanced Users)
Not supported by your current membership plan. Order Now.
HTTP Basic Authentication
Username:
Password:
Form Submission
Form Variables:
Cookie Support
Cookie Data:
Content Filters (Advanced Users)
Not supported by your current membership plan. Order Now.
Regular Expression:
Notification Options (Advanced Users)
Silent:
Send cd-diff file:
Error Handling:
Default
Notification Triggers (Advanced Users)
if the page size
changes by at least "x"
Internet


Edition Affichage Favoris Outils ?
Adresse
http://www.changedetect.com/changedetect/member/view_monitor.asp?mt_idx=169763&requestid=modify
OK
Précédente
Rechercher
Favoris
Google
Hotmail
QuickPost
Flashmail
Norton Internet Security
Copernic Agent
Le Web
Notification Options (Advanced Users)
Silent:
Send cd-diff file:
Error Handling:
Default
Notification Triggers (Advanced Users)
if the page size changes by at least "x" bytes:
when all of the words are detected:
when the exact phrase is detected:
when at least one of the words is detected:
when one of the words is not found:
Submit
Reset
Cancel
Terms of Service | Privacy Policy | Permission Policy
AdSpace: credit cards | credit cards | home loans | search engine optimization | free hosting | clickbank rss script | adsense psa alternative | online shop | detox
Copyright © ChangeDetect 2002-2005
Page Last Updated: November 09, 2004
Internet

Les newsletters

- Agent de push relativement efficace
- Listes de diffusion gratuite souvent
- Parfois « commerciale »
- De l'information souvent redondante
- Envoyée quelle que soit l'information
- Utilisation d'email anonyme recommandée

Google Alerts

- Surveillance dans les actualités de Google
- Gestion des alertes par création de compte
- Envoi par mail quotidien, hebdomadaire, selon l'actualité
- Un outil efficace pour surveiller la Presse Quotidienne Nationale et les sites d'actualités
- Peu efficace pour la PQR

Gestion des bookmarks

- Objectif basique
  - Retrouver une information, une société, un service,un produit
- Objectifs avancés
  - Trouver des nouveaux sites correspondant à un besoin informationnel
  - Accéder à ses bookmarks n'importe où, Pouvoir partager ses favoris
- Fonctionnalités
  - Stocker, organiser, annoter
  - Importer / Exporter
  - Surveillance de favoris
  - Catégorisation automatique

Yoono

- Logiciel de gestion des favoris + lecteur RSS + création et diffusion de flux RSS – version béta
- Proposition de favoris « proches »
- Diffusion en ligne de ses favoris
- Abonnement aux favoris d'autres utilisateurs
- Synchronisation des favoris sur serveur (gestion de dossiers privés / par défaut pas de partage des favoris stockés)
- Alerte sur nouveaux favoris identifiés
- A tester aussi : Human Links

Aspirateur de site (1)

- Initialement : limiter l'utilisation de la bande passante et les coûts
- Désormais
  - Permettre la consultation d'un site en mobilité
  - Trie les ressources (pages, doc pdf,mail…)
  - Peut permettre l'intégration dans un corpus de document plus global à traiter (text-mining)

Aspirateur de site (2)

- Installation d'un logiciel
- Définition de la requête
  - Récupération de l'adresse à surveiller
  - Définition de la profondeur de crawl
  - Choix des documents à récupérer
- Capture en local
- Possibilité de mise à jour de la requête
  - Certains proposent une mise en évidence des changements
    - http://www.wysigot.com/fr/
    - Check&Get : http://activeurls.com/en/
    - Memoweb : http://www.goto.fr/memoweb/index.asp

Paramétrage

- Déterminer la page de départ
- Niveau d'exploration des liens internes et externes
- Type de documents récupérés
- Nommage et organisation de la capture
- Paramétrage d'une requête sur Memoweb

Outil de vérification de lien

- Permet de s'assurer de la validité des liens
  - Sur des bookmarks
  - Sur un site web
  - Sur des pages surveillées
    - Nettoyage de bookmarks
    - Identification de bugs ou d'erreur de saisie en back office de site web
    - Réactualisation de contenu
- Démonstration sur Weblink Validator

Statistiques webs

- Permettait d'identifier les erreurs, les évolutions coté client, la charge serveur
- Permet de connaître
  - Sa zone géographique
  - Ses visiteurs
  - Les documents les plus utilisés
  - Ses mots clés d'entrée
    - http://www.bcarayon-ie.com/stats/

Outils commerciaux

- Livres blancs
  - Document méthodologique
  - Fortement orienté et peu objectifs
  - Véritable outil de promotion pour certains professionnels
- Newsletter
  - Affirmer sa compétence
  - Récupérer des contacts
  - Favoriser l'échange : on donne pour recevoir