Guide de recensement des outils de collecte, de traitement et de visualisation de l'information

Janvier 2006

CIGREF

« PROMOUVOIR L'USAGE DES SYSTEMES D'INFORMATION
COMME FACTEUR DE CREATION DE VALEUR POUR

Les outils de collecte, de traitement et de visualisation de l'information constituent un support essentiel de toute démarche d'Intelligence Economique d'Entreprise. L'industrie logicielle française est composée de nombreux acteurs industriels et laboratoires dont les technologies, performantes et souvent innovantes, demeurent, pour le moment, mal connues.

Afin d'identifier ces outils et ainsi donner aux DSI et autres directions métiers une meilleure visibilité de l'état du marché, le Cercle d'Intelligence Economique du CIGREF a souhaité diffuser ce document de synthèse, fruit d'une étroite collaboration avec la mission du Haut responsable à l'Intelligence Economique et la DCSSI au Secrétariat Général de la Défense Nationale.

Avertissement :

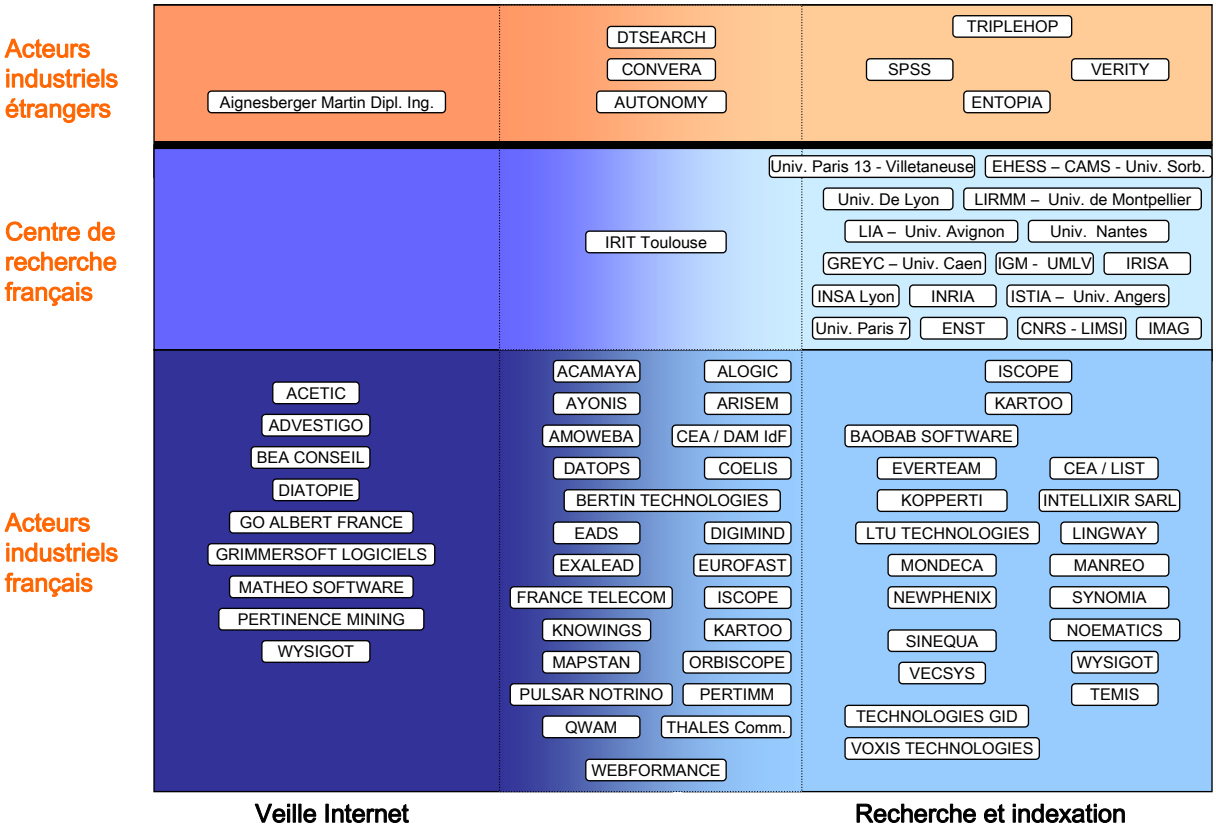
Ce premier recensement, établi courant 2005, n'est certainement pas complet. Certains détails et acteurs peuvent avoir été omis. Etant appelé à évoluer et être enrichi, ces fiches seront actualisées au fur et à mesure. Nous vous invitons à nous faire part de vos remarques et commentaires.

Tableau de synthèse

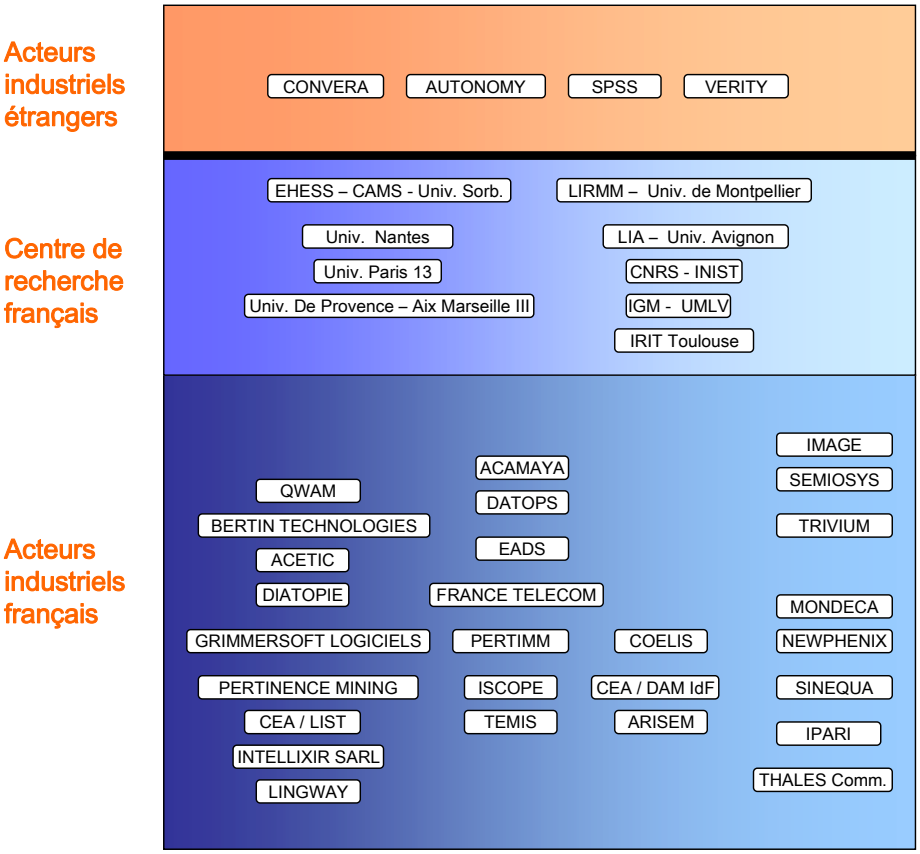
	Veille Internet	Recherche et indexation	Text mining	Data mining	Traduction	Traitement de l'image	Traitement de La parole	Représentation graphique	Knowledge management
1ère partie : Recensement des acteurs industriels français									
ACAMAYA	X	X	X						X
ACETIC	X		X						
ADVESTIGO	X								
ALOGIC	X	X							X
AMOWEBA	X	X						X	X
ARISEM	X	X	X						
AYONIS	X	X						X	
BAOBAB SOFTWARE		X							
BEA CONSEIL	X								
BERTIN TECHNOLOGIES	X	X	X	X				X	
CASTELIS									X
CEA / DAM Ile de France	X	X	X	X				X	
CEA / LIST		X	X	X					
COELIS	X	X	X	X					
Groupe DATOPS	X	X	X	X				X	
DIATOPIE	X		X					X	
DIGIMIND	X	X						X	
EADS (Defense and Communications Systems)	X	X	X	X	X	X		X	
ELIKYA									X
EUROFAST	X	X							
EVERTEAM		X							
EXALEAD	X	X	X		X				
FRANCE TELECOM R&D						X			
FRANCE TELECOM	X	X	X					X	X
GO ALBERT FRANCE	X								
GRIMMERSOFT Logiciels	X		X	X				X	
IMAGE			X						
Iparl			X	X				X	
INTELLIXIR SARL		X	X	X					
ISCOPE	X	X	X	X				X	
KARTOO	X	X						X	
KNOWINGS	X	X							X
KOPPERTI		X						X	
LINGWAY		X	X		X				
LTU TECHNOLOGIES		X				X			
MANREO		X				X			

MAPSTAN	×	×						×	
MATHEO SOFTWARE	×			×				×	
MONDECA		×	×						×
NEWPHENIX		×	×			×			
NOEMATICS		×							
NOHETO									
ORBISCOPE	×	×							
PERTIMM	×	×	×						
PERTINENCE MINING	×		×						
PULSAR NOTRINO	×	×						×	
QWAM	×	×	×	×					
SEMIOSYS			×	×				×	
SINEQUA		×	×						
SOFTISSIMO					×				
SYNOMIA		×							
SYSTRAN					×				
TECHNOLOGIES - GID		×							
TEMIS		×	×						
THALES Communications	×	×	×	×		×		×	×
TRIVIUM			×					×	×
VECSYS		×					×		
VOXIS TECHNOLOGIES		×					×		
WEBFORMANCE	×	×							
WYSIGOT	×								
XYLEME	×	×							

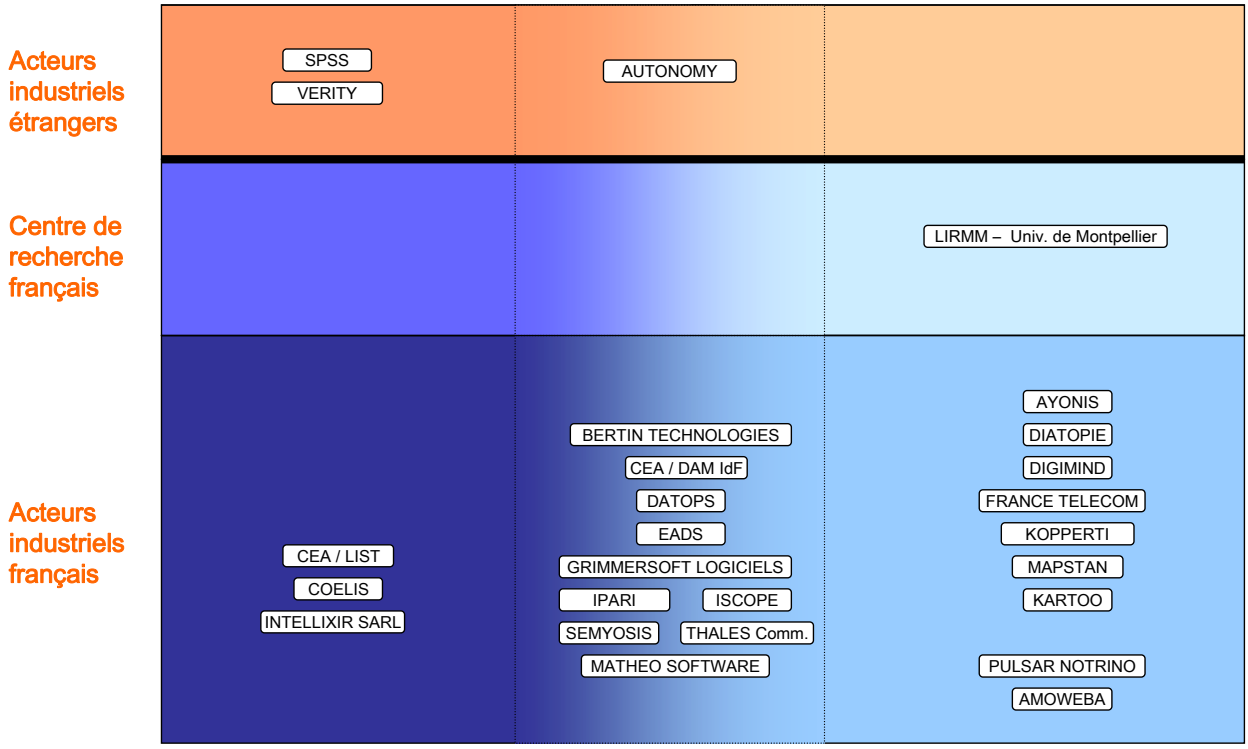
1. Veille Internet – Recherche et indexation



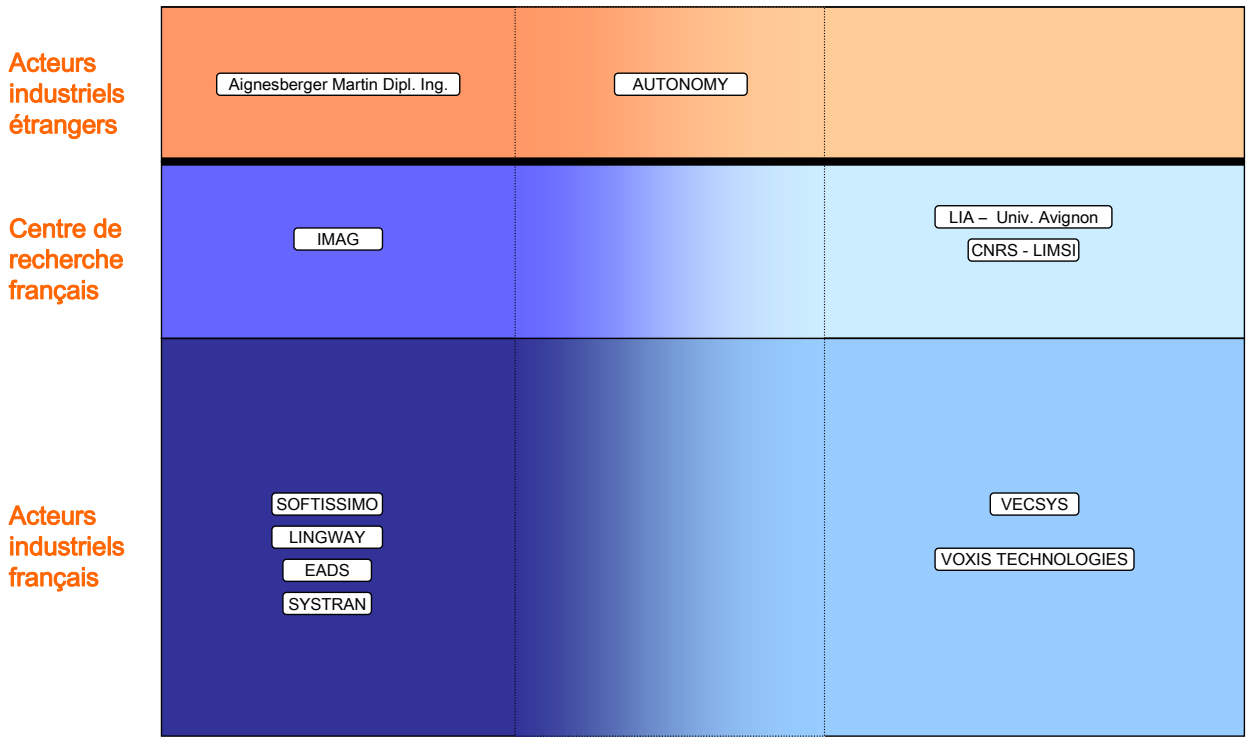
2. Text Mining



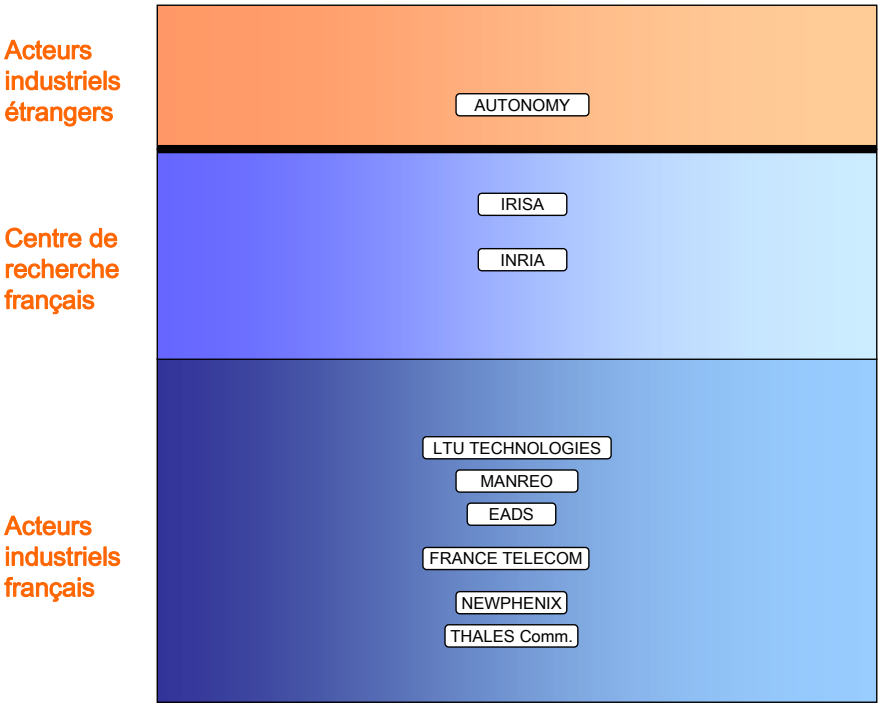
3. Data Mining - Représentation graphique



4. Traduction – Traitement de la parole



5. Traitement de l'image



1ère partie : Recensement des acteurs industriels français

ACETIC **(**ACTION ETUDE INFORMATION COMMUNICATION) www.acetic.fr	• **veille Internet** • recherche et indexation • **text mining** • data mining • traduction • traitement de l'image • représentation graphique • knowledge management	

Adresse :
45 rue Saint-Sébastien 75011 PARIS

Date de création : 1994

Partenariats :
GFI (revendeurs, co-développeurs, ..), ACCOM

Concurrents :
Lingway

Produits :
Analyse de textes ou de documents, moteur de recherche sémantique, structuration automatique de contenus et traitements d'enquêtes sociologiques.
Logiciel phare: Trope Zoom (permet l'index de documents).

Seule entreprise a proposé l'analyse du discours (explication de texte).

Brevets pour la structuration, la linguistique et l'ergonomie.
Récompensée par l'ANVAR (+580% CA en 5 ans)

ADVESTIGO
www.advestigo.com

- **veille Internet**
- recherche et indexation
- **text mining**
- data mining
- traduction
- traitement de l'image
- représentation graphique
- knowledge management

Adresse :
1 Rue Royale,
Bât. D, 9ème étage
92213 Saint-Cloud,
France

e-mail : advestigo@advestigo.com
téléphone : 01 72 77 7000
fax : 01 46 89 68 60

Date de création : 2002

Partenariats :
Soutien : fonds de capital-risque de premier plan et par des partenaires institutionnels tels que l'ANVAR, Scientipôle-Initiative ou l'ANRT.
Concurrents **:**

Produits :

Créée en octobre 2002 par le Dr Hassane Essafi et le Dr Marc-Michel Pic, tous deux issus de la recherche publique au CEA-LETI.

Spécialisée en « Protection d'Actifs Numériques ".

Technologie : calcul d'empreintes numériques (la " théraographie ") permet de reconnaître des copies exactes ou approchées, totales ou partielles d'un contenu original. Documents concernés : texte, son, image, vidéo, animations ou code structuré et ne nécessite aucun marquage préalable du document d'origine.

Technologie basée sur l'extraction d'empreintes de contenus multimédia variés.

C'est une technologie innovante car elle ne nécessite :

- Pas de tatouage (" watermarking ")
- Pas d'encapsulation
- Pas de signature électronique

AdvestiSEARCH vous permet de protéger vos actifs numériques par une surveillance continue et automatisée des protocoles de communication (Internet, P2P, Newsgroup…), par un contrôle à façon de l'utilisation suspecte des contenus et par une mesure précise de leur diffusion.

ALOGIC www.alogic.fr	• **veille Internet** • **recherche et indexation** • text mining • data mining • traduction • traitement de l'image • représentation graphique • **knowledge management**	

Adresse :
4, rue Galilée
75116 Paris
Tel : (33)1 53 17 53 17
Fax : (33)1 47 23 65 19

Contact : contact@alogic.fr

Date de création : Mai 1999

Partenariats : Pertimm, Oracle, Coframi **Concurrents** : Autonomy, Arisem, Convera		

Produits :
ALOGIC propose une offre logicielle modulaire qui représente le cœur de sa technologie sémantique: "aperto libro" (surveillance Web, moteur de recherche et d'indexation pour tous formats de documents, outils de classification, de résumé et traitements linguistiques).

AMOWEBA (racheté par Social Computing) **www.amoweba.com**	• **veille Internet** • **recherche et indexation** • text mining • data mining • traduction • traitement de l'image • **représentation graphique** • **knowledge management**	

Adresse :
89 Boulevard de Sébastopol, 75002 PARIS

Date de création : Août 2000

Concurrents :
Groove Network, Autonomy, Vista Portal,Neunet Solutions

Produits :
Mapstan et HumanLinks sont à la croisée de la veille, du KM et de la représentation graphique.

Mapstan : cartographie dynamique.
L'information est représentée sous forme d'entités reliées entre elles par des liens. Ces liens peuvent être fonctionnels, capitalistiques, géographiques, de compétences. Ce produit est accessible via un navigateur WEB. Le produit fonctionnant sur le serveur WEB cherche son information sur des bases de données.

MapStan Search : est une solution de recherche étendue fournissant une représentation cartographique des résultats d'une requête. Il intègre un processus d'analyse unique et un archivage systématique de tous les résultats. Ainsi, MapStan Search enrichit la représentation des résultats de recherche en recommandant des informations issues de l'analyse de l'ensemble des requêtes déjà réalisées. Ainsi chaque nouvelle recherche est capitalisée et accroît la pertinence des réponses fournies.

HumanLinks : client de veille collaborative.
Ce produit se veut être le guichet unique d'accès, de traitement et de partage de l'information distribuée. Il est composé d'un programme client (installé sur chaque poste de travail) qui interroge toute sorte d'information. L'indexation et la classification des informations se fait par calcul et non pas par approche sémantique ou linguistique.
Les résultats sont traduits sous forme graphique, ce qui permet de classer intuitivement les informations. Ils sont ensuite partagés entre les membres d'un même groupe.

Services :
Etudes d'usages, bancs de test.

ARISEM www.arisem.com	• **veille Internet** • **recherche et indexation** • **text mining** • data mining • traduction • traitement de l'image • représentation graphique • knowledge management	

Adresse :
1 Av Carnot 91300 Massy

Date de création : 1996

Partenariats :
SINEQUA

Concurrents :
VERITY, AUTONOMY, DIGIMIND

Produits :
L'outil d'ARISEM est un crawler (outil d'aspiration de sites Internet) multilingue (français, italien, anglais, espagnol, allemand), déclinable en version monoposte ou serveur.
Il nécessite la constitution et l'administration d'un référentiel métier.

AYONIS www.ayonis.com	• **veille Internet** • **recherche et indexation** • text mining • data mining • traduction • traitement de l'image • **représentation graphique** • knowledge management	

Adresse :
Parc d'affaires des Portes, rue Sainte Marguerite 27100 Val de Reuil

Date de création : 1993

Partenariats :
Groupe Autom-Tech

Concurrents :
KB Crawl

Produits :

VS 2000 : outil de veille dédié aux PME/PMI
Il est composé de cinq modules :
VS ASPI (surveillance de sites, newsgroups et forums), VS Admin (paramétrage des profils utilisateurs), VS Serveurs (traitement et diffusion de l'information), VS Biblio (analyse et cartographie de l'information), VS Search (outil d'indéxation, et de recherche bilingue)

BAOBAB Software www.baobab-software.fr	• veille internet • **recherche et indexation** • text mining • data mining • traduction • traitement de l'image • représentation graphique • knowledge management	

Adresse
66, rue Michel Ange, 75016 PARIS

Date de création : 11/1998

Produits :
DOC Accelerator, est un moteur de recherche par balayage de documents texte (format word, pdf, rtf, MS office, Notepad) ou de mails (outlook express, exchange, lotus notes) permettant de retrouver très rapidement (moins d'une seconde) par un système de « mot-clé » un document à travers des milliers de mails ou de pièces jointes avec la certitude d'avoir une information complète et fiable, en éliminant les documents sans intérêt. La différence avec les produits existants réside essentiellement dans l'absence de « thésaurus » (sorte de dictionnaire), mais repose sur un algorithme simple très facile d'installation.

BEA CONSEIL
BRUNO ETIENNE ET ASSOCIES
www.beaconseil.com

- **veille Internet**
- recherche et indexation
- text mining
- data mining
- traduction
- traitement de l'image
- représentation graphique
- knowledge management

Adresse :
10, rue Lionel Terray, 92508 Rueil-Malmaison

Date de création : 27/07/1995

Partenariats :
ANVAR, FLA Consultant, ADIT

Concurrents :
Iscope, ARISEM

Produits :
KB Crawl : surveillance de sites Internet.
Envoi de mél lors de chaque modification ou mise à jour d'URL surveillées. L'interface utilisateur permet ensuite de naviguer dans les différentes versions d'un site. Des fonctionnalités permettent également la récupération de données, de filtrage et de paramétrage des formulaires permettant de programmer tout type de surveillance de sites web:
- interfaçage avec des moteurs d'analyse de données (text minig),
- filtrage avancé avec des URL téléchargées,
- gestion automatique des formulaires cachés,
- gestion des variables de sessions.
Possibilité de tester KB CRAWL gratuitement pendant trente jours.

BERTIN TECHNOLOGIES www.bertin.fr	• veille Internet • **recherche et indexation** • **text mining** • **data mining** • **traduction** • traitement de l'image • **traitement de la parole** • **représentation graphique** • knowledge management	

Adresse :
Parc d'activité du Pas du Lac
10 bis av Ampère
78180 Montigny-le-Bretonneux

Date de création : Janv 2000

Partenariats :
Dans le domaine du traitement automatique des langues : Vecsys, LIMSI, CEA

Produits :
Indexal : moteur de recherche et de veille

Bertin Technologie s'appuie sur les compétences spécialisées de Vecsys, du LIMSI et du CEA/LIST : Reconnaissance de la parole, reconnaissance de thèmes crosslingue et de l'image.

CASTELIS www.castelis.com	• veille Internet • recherche et indexation • text mining • data mining • traduction • traitement de l'image • représentation graphique • **knowledge management**	

Adresse :
11 rue Maurice GRANCOING, 94200 IVRY SUR SEINE

Date de création : 2000

Partenariats :
CEGID, MICROSOFT, SAGE, DELL

Produits :
Edition de progiciels. Intégration de progiciels de gestion. Développement d'outils de finance, gestion, comptabilité en technologie Intranet. Analyse, révision et reprise de données. Réalisation de solutions spécifiques.

ATLAS : logiciel répondant aux besoins de gestion documentaire et de partage des connaissances.
- Indexation en temps réel de tout type de documents
- Administration autonome et intégrée
- Création d'actualités
- Moteur de recherche
- Forums
- Adaptation à chaque charte graphique
- Navigation simplifiée
- Génération automatique d'un plan du site
- Statistiques de connexion

Technologies employées : Transac SQL, HTML/ASP, JavaScript, OCX

CEA / DAM Ile de France DSSI : Département des Sciences et Simulation de l'Information CDEI : Centre de Documentation et d'Exploitation de l'Information	• **veille Internet** • **recherche et indexation** • **text mining** • **data mining** • traduction • traitement de l'image • **représentation graphique** • knowledge management	

Adresse :
BP n°12 – 91680 Bruyère le Châtel

Partenariats :
Partenariat limité au CEA en particulier CEA/LIST.

Produits :
Pour l'instant, le CDEI n'a pas de produit prêt à la vente.
Il est co-titulaire d'un brevet avec J.F. Delpech, inventeur (US CREST) d'un outil de recherche documentaire par distance textuelle SIAS : pour cela, leur participation à ce brevet est de rechercher un marché et de commercialiser le produit.

Services :
Leurs prestations sont principalement centrées sur de l'assistance à maîtrise d'ouvrage et à la maîtrise d'œuvre en mettant leur expertise d'ingénierie documentaire au service de leur client.
Cette expertise est centrée sur les technologies et les outils de :

- Aspiration de site,
- Moteur de recherche,
- Textmining,
- Datamining,
- Cartographie sémantique,
- Analyse statistique,
- Repository.

CEA / LIST
Laboratoire d'Ingénierie de la Connaissance Multimédia Multilingue (LIC2M)
www-drt.cea.fr

- veille Internet
- **recherche et indexation**
- **text mining**
- **data mining**
- traduction
- traitement de l'image
- représentation graphique
- knowledge management

Adresse :
BP n°6 – 92265 Fontenay aux Roses Cedex

Date de création : janvier 2002

Produits :

XEDIX : base de données XML native de hautes performances qui assure les fonctionnalités suivantes :

- Stocker, gérer et indexer des ressources XML ;
- Assurer l'indexation des ressources à n'importe quel niveau de granularité ;
- Traiter des requêtes complexes grâce à son langage permettant l'adressage dans les arbres XML selon des constructions Xpath ;
- Stocker et gérer des images résultant de la numérisation de documents originaux et de leur représentation textuelle partielle ou complète. Les images sont affichées parallèlement aux ressources textuelles retrouvées par les requêtes XML. Les mêmes possibilités sont offertes pour des vidéos ou toute autre ressource non XML associée à des métadonnées XML,
- Gérer les utilisateurs et les groupes d'utilisateurs, leur allouer les droits d'accès aux ressources individuelles ou à des groupes de ressources. Les droits peuvent être définis à tout niveau de granularité, de l'instance XML complète jusqu'aux éléments feuilles.

Technologies : C++ avec utilisation des STL, HTML et PHP (V.4).
O.S. : UNIX et Linux (V2.2 et plus), Solaris 8. Il peut être compilé pour d'autres systèmes POSIX.

PIRIA (programme d'indexation et de recherche d'images par affinité) :
Ce produit permet de :

- Indexer les images et les vidéos,
- Segmenter les vidéos,
- Rechercher et inversement classer par similarité,
- Traiter les requêtes textuelles et/ou image,
- Fusionner les recherches.

COELIS http://www.coelis.com,	• **veille Internet** • **recherche et indexation** • **text mining** • **data mining** • traduction • traitement de l'image • représentation graphique • knowledge management	

Adresse :
8 rue du marché, 67 000 Strasbourg

Date de création : 2001

Partenariats :
IMEXPERT

Produits :
Coexpect : Moteur de recherche et surveillance de site - robot paramétré en fonction des sources à interroger.
Coevista : Traitement sémantique et filtrage par apprentissage - composé d'un moteur de recherche basé sur une approche connexioniste (réseau de neurones) et d'agents intelligents.
Coevision : Traitement, recoupement et cartographie -
Coesia : Moteur de recherche, surveillance de site, text et data mining (Intègre Coexpect, Coevista et Coevision)

Groupe DATOPS

DATOPS CONSULTING – DATOPS SA.
http://www.datops.com

- **veille Internet**
- **recherche et indexation**
- **text mining**
- **data mining**
- traduction
- traitement de l'image
- **traitement de la parole**
- **représentation graphique**
- knowledge management

Adresse :
60, rue de Ponthieu 75008 Paris
Parc Georges Besse, Allée Charles Babbage30000 Nîmes
Bureaux à Londres et New-York

Date de création :
- le 01/07/1994 pour Datops Consulting (anciennement CMC puis Startem)
- le 28/11/1994 pour Datops SA devenu en 2001 le groupe DATOPS

Partenariats :
France Telecom (pour l'hébergement et la sécurisation de leur solution), CNRS (laboratoire de linguistique GREYC) de Caen.

Concurrents :
DIGIMIND, ARISEM, CYBION.

Activité :
Le Groupe Datops est spécialisé dans le risque informationnel, la veille, l'analyse de l'information et l'intelligence économique. Cette société regroupe en son sein un éditeur de logiciel (Datops SA) qui développe notamment le progiciel Péricles et les solutions InfoMonitor et InfoMetrix et une société de consulting (Datops Consulting), spécialisée dans le conseil pour la mise en place de dispositifs de veille et d'anticipation et de maîtrise des risques business.

Produits :
Périclès Veille : logiciel de veille (crawling et rapatriement d'informations à partir de sources électroniques : base de données, presse, newsgroups, forums, sites institutionnels, etc.).
Periclès repose sur des agents automatiques d'extraction paramétrés par un opérateur humain et sur un module Wise de pondération de l'information. Des index lexicaux et linguistiques ainsi que des indicateurs de tonalité des données (positif, négatif, neutre) sont utilisés dans une 2ème phase avec une action de thématisation des données.
Périclès Analyse : représente de façon graphique les volumes et les tendances de l'information, l'origine géographique de l'information, les auteurs les plus actifs, les thèmes abordés, etc…
Il comprend différents modules : **InfoMonitor**, portail de veille et d'alerte, **InfoMetrix** pour l'évaluation sous forme de tableaux de bord et graphiques et **RiskMetrix,** produit de mesure du risque conçu pour l'assurance.

DIATOPIE www.diatopie.com	• **veille Internet** • recherche et indexation • **text mining** • data mining • traduction • traitement de l'image • **représentation graphique** • knowledge management	

Adresse :
27 Boulevard St Martin 75003 PARIS.

Date de création : 1998

Partenariats :
Alain LELU – Université de Franche Conté
Noematics (pour le moteur d'indexation)

Concurrents :
ARISEM, LEXI, SAS

Produits :
Neuronav : outil de text mining consacré à l'analyse et au parcours de corpus textuels. Il facilite l'extraction de documents pertinents dans un corpus documentaire volumineux (jusqu'à 200.000 documents). Le moteur d'indexation est celui qui est développé par la société Noematics. Le grand atout du logiciel réside dans la cartographie sémantique : Cartoweb.

Atouts technologiques :
Utilise le moteur d'indexation Réflexion de Noematics.

DIGIMIND www.digimind.fr	• **veille Internet** • **recherche et indexation** • text mining • data mining • traduction • traitement de l'image • **représentation graphique** • knowledge management	

Adresse
Siège : 12, rue Ampère -BP 267-38016 Grenoble cedex 01
Bureau commercial : 32, rue de paradis, 75010 Paris

Date de création : 1998

Partenariats :
Pas de partenariats réguliers, DIGIMIND fait appel de manière ponctuelle à des consultants.
HM et associés, Sesame, Competia, CCIP, Mediapps.

Concurrents :
Concurrents :
Français : Arisem, Datops
Etrangers : Cipher Systems, LLC (US), Comintell AB (Suède), Novintel (finno-canadien), Strategy Software, Inc (US), Traction Software, Inc. (US, In-Q-Tel), Wincite Systems, LLC (US), Autonomy, plc (UK).

Produit :
Digimind Evolution V 1.8 :
Progiciel intégré de veille stratégique permettant d'assurer l'ensemble des fonctionnalités nécessaires à un dispositif de veille stratégique :
- Sourcing : gestion collaborative de bookmarks multisources (web, forums, newsletters, …), connexion aux moteurs de recherche (Web Invisible) et aux bases de données (Factiva, Dalog, Jane's,…),
- Surveillance : agents intelligents distribués identifiant, extrayant et agrégeant toute nouveauté apparaissant sur les sources surveillées et correspondant aux requêtes des veilleurs.
- Analyse : espace collaboratif multi-utilisateurs permettant l'édition, l'enrichissement, le classement, le croisement, et l'analyse à l'aide entre autre d'outils cartographiques, des informations collectées, par les personnes autorisées.
- Diffusion : générateurs semi-automatique ou automatique de livrables tels que rapports de veille, fiches concurrents, tableaux de benchmark, cartes d'acteurs, etc…et diffusion multicanaux aux décideurs (versions imprimables, pushmail, newsletter, portail dédié, push XML vers Intranet,…)

Il repose sur l'algorithme « iSCRAP » (développé par M EL HADDAR en février 2004 mais non encore protégé par brevet), qui permet une extraction automatisée d'actualités en ligne pour la veille stratégique.

Monitor : Surveillance de pages et de sites Internet. Fonctionne uniquement en mode ASP - Connexion au site de Digimind.

Strategic Finder : Méta-moteur.

EADS (Defense and Communications Systems) Pôle Defense and Security Systems www.eads.net	• **veille Internet** • **recherche et indexation** • **text mining** • **data mining** • **traduction** • **traitement de l'image** • **représentation graphique** • knowledge management	

Adresse :
6, rue Devoitine – BP14 – 781423 Vélizy Villacoublay Cedex

Partenariats :
INRIA Nancy (LORIA) – INRIA Rouen – Université de Rouen – Laboratoire Informatique du Havre – Université de Paris 6 (LIP6) – LIH – LUCID'IT – GERDOSS – CEA – LaRIA – LIPN – PME/PMI – Réseaux d'experts et projets de recherche communs au sein du groupe.

Concurrents :
THALES – CEA – France Telecom.

Produits :
Une plate-forme OSINT « WEBLAB » est en préparation.

X-Miner : plat-forme multicomposant pour la structuration dynamique de documents en texte libre (outils d'acquisition et d'analyse de l'information) :
- Agents de surveillance de pages web,
- Aspiration,
- Structuration et analyse du contenu,
- Alerte par push.

P-Miner : outil de veille économique sur Internet :
- Scénarios d'aspirations (pages statiques, dynamiques et formulaires),
- Veille programmable pour départ différé,
- Suivi des aspirations (log et statistiques),
- Structuration et analyse du contenu.

NetProtect 1 et 2 : Systèmes de classification automatique de pages web.

ACTD : Prototype d'une plate-forme de recherche et d'exploitation de sources ouvertes pour la simulation.

EVERTEAM www.ever-team.com	• veille Internet • **recherche et indexation** • text mining • data mining • traduction • traitement de l'image • représentation graphique • knowledge management	

Adresse :
40B rue de la Vilette 69003 Lyon 3
Bureau : 4 place Felix Eboue 75012 Paris

Date de création : 1977

Partenariats :
Steria

Produits :
Moteur de recherche – catégorisation
Autres composantes : GED - WORKFLOW

EXALEAD www.exalead.com	• **veille Internet** • **recherche et indexation** • **text mining** • data mining • **traduction** • traitement de l'image • représentation graphique • knowledge management	

Adresse :
10, place de la Madeleine
75008 PARIS
Bureaux : Italie, USA
Date de création : 09/2000

Produits :
Exalead one:search 4.0 : première plate-forme de recherche unifiée sur les PCs, les serveurs d'entreprise et le Web.
Exalead one:search est au cœur de tous les logiciels de la gamme exalead :

- Exalead one:desktop : outil de recherche sur PC (disques durs et messageries : Microsoft Outlook, Microsoft Exchange et Lotus Notes), téléchargeable en version gratuite et professionnelle sur exalead.com
- Exalead one:workgroup : indexation des serveurs Windows, téléchargeable sur exalead.com dès février 2006
- Exalead one:enterprise, logiciel d'indexation haute performance et temps réel, capable de traiter tout type d'information : web, bases de données, intranets, système de gestion documentaire, messagerie, etc. Exalead one:enterprise dispose d'outils d'alertes.
- Exalead one:datacenter, destiné aux applications industrielles à forte volumétrie et/ou fort trafic. Deux applications existantes :
  - AOL.fr (depuis 2002) : 80 millions de pages Web et plusieurs millions de requêtes / jour
  - exalead.com : 4 milliards de pages indexées en janvier 2006. Il peut être couplé à des applications spécifiques de type veille développables par Exalead ou un partenaire.

Exalead one:search permet l'accès aux sources en temps réel et réalise l'indexation immédiate d'un document dès qu'un changement est opéré.
Exalead one:search est initialement présenté avec des interfaces en français, anglais, allemand et chinois, des interfaces dans d'autres langues (nécessitant très peu de travail) seront disponibles ultérieurement (en commençant par les langues européennes et quelques langues asiatiques).

Atouts technologiques :

- Technologie développée depuis 1999.
- Indexation jusqu'à plusieurs milliards de documents
- Performances :
  - 20 millions de documents par serveur (un biprocesseur supporte l'équivalent de 100 000 utilisateurs).
  - 100 millions de pages Web par serveur
- L'index Exalead se comporte comme une base de données XML et comporte des champs textuel, numérique (tri), date, etc. ainsi que des champs « arborescents » dont il est possible de faire la synthèse et qui peuvent être utilisés pour naviguer.
- Dédoublonnage.
- Traitement du Web invisible (pages non indexées, accès aux ressources situées à des emplacements non standards, traitement des pages dynamiques, suivi des liens JavaScript).
- Langues traitées : Près de 50 langues traitées, une quinzaine avec des outils avancés comme la lemmatisation ou la génération automatique de mots-clés (Anglais, Allemand, Français, Suédois, Espagnol, Italien, Hollandais, Portugais, Norvégien, Finnois, Danois, Chinois, Arabe, Japonais, etc.).

- Large gamme de connecteurs (http, file system, ODBC, Lotus Notes, Documentum, Livelink, etc.)
- Cross-linguisme réalisé au travers de partenariats (Lingway, Systran)
- Outils de catégorisation automatique à forte capacité à monter en charge :
    - Classification dynamique (clustering) par extraction de terminologie (phrases-clé).
    - Extraction d'entités nommées par un système de transducteurs morpho-syntaxique basé sur XML et des règles métier.
    - Sélection des phrases pertinentes.
- Fonctionnalités automatiques (et sans dictionnaire) de recherche phonétique et approchée (entièrement paramétrable par règles) et de correction orthographique.
- Opérateurs de recherche avancée complexes dont recherche par motifs (expressions régulières arbitraires) et opérateur « NEAR »
- Architecture ouverte : SDK, APIs JAVA et XML et Portlets
- Délais de mise en œuvre réduits

FACTIVA www.factiva.com/fr	• **veille Internet** • **recherche et indexation** • **text mining** • data mining • traduction • traitement de l'image • représentation graphique • **knowledge management**	

Adresse : 6-8 Bd Haussmann 75009 Paris

Date de création : 1999

Partenariats : Microsoft, IBM, Verity, Salesforce.com

Concurrents : Dialog, Presse ED, Lexis Nexis

Produits :

Factiva.com
Factiva Insight : Reputation Intelligence
Factiva Insight : Media Intelligence
Factiva SalesWorks
Factiva Public Figures and Associates
Factiva Select
Factiva Track module
Factiva Publisher

FRANCE TELECOM R&D

Branche Développement FT/R&D/DIH/HDM
DIH = Direction Interaction Humaine
HDM = Hyperlangage et Dialogue Multimédia
www.francetelecom.com/rd

- veille Internet
- recherche et indexation
- text mining
- data mining
- traduction
- **traitement de l'image**
- représentation graphique
- knowledge management

Adresse :
4, rue du Clos Courtel – BP 59 – 35512 Cesson Sévigné

Partenariats :
MANREO (la base de données commune entre MANREO et FT, intégration par FT dans STUDIO CREATIF de Hypercast Editor, Hypercast Publishing Server et Hypercast Warehouse de MANREO).

Concurrents :
THALES – EADS – CEA

Activité :

- Codage et compression des contenus audio
- Codage et compression de l'image animée
- Reconnaissance et indexation sonore :
  - Discrimination parole et musique
  - Reconnaissance de sons divers (en démarrage)
  - Séparation et reconnaissance de sources sonores (en démarrage)
  - Systèmes de reconnaissance et transcription de parole pour les contenus audiovisuels
- Reconnaissance et indexation vidéo
  - Identification de différentes transitions vidéos (cuts, fondus enchaînés)
  - Détection et reconnaissance de visages
  - Détection de textes (incrusté, de scène)
  - Détection d'objets génériques
  - Similarité visuelle et classification
- Insertion de messages indétectables dans une bande sonore ou une bande vidéo
- Traitement automatique du language

Produits :
France Télécom développe en parallèle deux types de produits : les algorithmes et l'atelier.

- Algorithme de reconnaissance de la parole pour les applications d'indexation audiovisuelle – solution d'indexation phonétique multi-locuteurs pour les contenus vidéos

- Projet STUDIO CREATIF : plate-forme d'indexation et de gestion de vidéos intégrant :
  - Codage vidéo
  - Moteur d'analyse texte – parole – son – image - vidéo
  - Génération automatique de résumés de vidéo
  - Segmentation vidéo en objet
  - Recherche d'images ou de régions similaires dans les images
  - Classification des images
  - Détection et reconnaissance de visages
  - Détection de texte dans les images

Cette plate-forme est conçu pour pouvoir intégrer les meilleures technologies du moment d'analyse des différents médias : AMIRAL (Architecture Modulaire d'Indexation et de Recherche Audiovisuelle en Ligne).

FRANCE TELECOM

SCE/DBI Business Développement

www.francetelecom.com

- **veille Internet**
- **recherche et indexation**
- **text mining**
- data mining
- traduction
- traitement de l'image
- **représentation graphique**
- **knowledge management**

Adresse
116 Avenue VERDIER, 92120 MONTROUGE

Date de création : 1997

Partenariats :
Factiva (broker d'informations), Softissimo (traduction), Leximine (textming et cartographie), Verity (moteur)

Concurrents :
Thales, Datops

Produits :
STATEASY : il s'agit d'un portail regroupant différents outils de veille. Les machines et les données sont externalisées au siège de France Télécom (mode ASP). L'accès au client se fait au travers d'un navigateur avec un chiffrement SSL.
Une étude est faite conjointement entre France Télécom et le client pour définir les filtres métier. Ensuite, l'utilisateur s'abonne et personnifie ses recherches et la présentation des données.
La facturation se fait au nombre d'utilisateurs (1.000 euros / an / utilisateur)

GO ALBERT FRANCE www.albert.com	• **veille Internet** • **recherche et indexation** • **text mining** • data mining • traduction • traitement de l'image • **représentation graphique** • **knowledge management**	

Adresse :
Siège social: 12 rue Vivienne – 75002 Paris
Bureau commercial : 19, rue du Quatre Septembre – 75002 Paris
R& D : Parc du Millénaire, 1025 Rue H. Becquerel 34000 Montpellier
Présent en Suisse et Angleterre via des revendeurs indépendants.

Date de création : 2001

Partenariats :
SPOTTER (veille sur la presse), BULL Services (marché de la Défense nationale), SQLI (intégration de systèmes), Xylene, COLT Télécommunications. En Suisse, MEGA Consultants (KM).

Concurrents :
France : Sinequa, Thales (Arisem-Kalima), Temis (partiellement), Exalead, etc.
Autonomy, Verity, Hummingbird, Convera

Produits :
AMI Enterprise Discovery (AMI ED) est une solution dédiée à la recherche fédérée d'informations au sein d'un Intranet d'entreprise. Permettant d'accéder aux différentes sources d'information, ses performances résident dans sa capacité à analyser les requêtes qui lui sont adressées, que celles-ci soient correctement orthographiées ou pas, et ce quelle que soit la langue utilisée, tout en se basant sur la connaissance du vocabulaire employé par l'utilisateur.

AMI Market Intelligence (AMI MI) est une application de veille individualisée et complètement personnalisable. L'utilisateur définit les thèmes ou sujets qui l'intéressent, éventuellement les sources d'information pertinentes (sites web, bases de données, NewsGroup), la fréquence de diffusion des résultats ainsi que la façon de les publier (par courrier électronique ou en alimentant une application propre à l'environnement client).

AMI Website Access : recherche fédérée d'information au sein du site Internet d'une entreprise. Il s'agit d'une solution d'assistance à la navigation facilitant l'accès au contenu du site, voire du web si l'information recherchée par l'internaute ne se trouve pas sur le site lui-même.
Ses fonctions :

- Recherche fédérée (connecteurs génériques et spécifiques),
- Analyse des questions et génération des hypothèses (gestion des synonymes, approximation phonétique et orthographique, expansion des requêtes),
- Analyse et restitution des résultats (fusion, trie par pertinence, regroupement par thèmes, format de restitution, confidentialité, fonctions d'apprentissage, pages préférentielles),
- Indexation – filtrage,
- Base de connaissances,
- Mémoire des utilisateurs.

Il inclut une fonction d'aspiration de sites externes et des fonctions linguistiques avancées (7 langues (UK, D, SP, I, NL, P et F), reconnaissance de la langue, gestion des mots vides, citations clefs).

Atouts technologiques :
AMI (Albert Meaning Interpreter) est une solution innovante de gestion des informations de l'entreprise étendue (« Extended Enterprise »). AMI permet de collecter, analyser et organiser les informations issues de sources disparates (bases documentaires Intranet ou extranet) afin de les rendre exploitables aux utilisateurs.

Les fondations du logiciel sont protégées par des **brevets** en Europe mais aussi aux Etats-Unis :
- « Grammaire Minimale Indépendante de la Langue »# B-3851
- « Enhancing online support » # B-3561
- « Interface en Langage Naturel » # B-3563
- E-commerce utilisant une Interface en Langage Naturel #B-3562

GRIMMERSOFT Logiciels
www.grimmersoft.com

- **veille Internet**
- recherche et indexation
- **text mining**
- **data mining**
- traduction
- traitement de l'image
- **représentation graphique**
- knowledge management

Adresse :
6 rue de Clignancourt 75018 Paris

Date de création : mai 1995

Produits :
WordMapper Text analysis : logiciel d'analyse textuelle et de cartographie de l'information, destiné à la veille technologique et à la veille sur Internet. (pages html des moteurs de recherche, articles des forums de discussion, e-mails, brevets, revues de presse, entretiens,...).

WordMapper Link Analysis : représente l'information structurée sous la forme de graphes en réseaux. Il adopte essentiellement une démarche statistique et graphique. La méthode consiste à construire des réseaux pour représenter les associations entre les modalités des variables à analyser (numéro, nom de personne,…).

QuestionData : traitement d'enquête.

StatBox : logiciel d'analyse statistique intégré à MS Excel et MS Access. Dédié aux spécialistes.

StatBox Pro : idem + modules spécialisés.

Atouts technologiques :
Maîtrise des algorithmes de représentation graphique de l'information structurée ou non structurée.

IMAGE www.image.cict.fr	• veille internet • recherche et indexation • **text mining** • data mining • traduction • traitement de l'image • représentation graphique • knowledge management	

Adresse :
12 rue Thiers, 31 400 TOULOUSE

Date de création : Juillet 1986

Partenariats :
CNRS, notamment pour la recherche fondamentale sur ALCESTE.

Produits :
ALCESTE : « analyse de données textuelles » pour le traitement d'enquêtes, analyse de discours, conseil en marketing, recherche documentaire…
L'objectif du logiciel est de quantifier un texte pour en extraire les structures signifiantes les plus fortes et d'en dégager l'information essentielle. La méthode utilisée repose sur l'analyse du vocabulaire, le découpage du « corpus » en unités de contexte, et la classification des unités en classes suivant la méthode de « classification descendante hiérarchique ». La version V5 permettra l'analyse de textes rédigés dans des langues étrangères.
La société développe également **IMAGEST** logiciel de gestion de données relatives à l'entreprise, et **STAT-EDUC**, logiciel de traitement de résultats d'enquête.

Iparl

Investigation par l'Image (I par I)
http://www.ipari.com
Groupe : I2

- veille Internet
- recherche et indexation
- **text mining**
- **data mining**
- traduction
- traitement de l'image
- **représentation graphique**
- knowledge management

Adresse :
48, rue des Vignoles / 75020 PARIS

Date de création : 19-04-1999

Partenariats :
Liens privilégiés avec :

- Société I2 (UK) pour les logiciels d'analyse visuelle (iBase, iBridge, Analyst's Notebook, ….).
- Société TOVEK (Tchéquie) pour les applications plein texte.
- Société S&A (Italie) pour les applications spécialisées téléphonie (TETRAS).
- AFP pour l'accès aux bases d'information.
- Société Speedware pour une base de données de technologie OLAP de gestion de gros volumes structurés (MICROLIS).

Produits :
Analyst's Notebook : Logiciel d'analyse visuelle spatiale, temporel. Comporte de fortes capacités d'analyse de recherche de chemins, de relations à plusieurs niveaux ou entre une période de temps, de recherche de réseaux, …

MAUD : utilitaire de mise en forme des données avant traitement par les autres logiciels de la solution.

TOVEK : Basé sur le moteur VERITY. Indexation de documents plein texte, recherche, pré analyse, et génération de graphes visuels de liens ou temporels des objets trouvés (mots ou groupes de mots) dans les documents originaux.

iBridge : Accès ergonomique à des bases de production de type ORACLE, SQL, ACCESS. iBridge n'extrait pas d'information de la base accédée, mais schématise les données existantes en permettant leur accès sous forme d'objets et de liens. Ces objets et liens permettent de poser des questions sans connaître la structure d'une base ni les langages de requête, et de représenter les résultats sous forme de graphe spatiaux ou temporels dans le logiciel Analyst's Notebook.

i2TextChart : Automatisation de la capture d'information non prédéfinies depuis des documents électroniques (Word, pdf , HTML, …). Mise en forme des objets et des liens (mots/concepts) trouvés. Envoi pour analyse et compilation vers le logiciel d'analyse visuelle Analyst's Notebook ou le gestionnaire de bases d'analyse iBase. I2TextChart est conforme à la démarche d'analyse et remplace le surligneur et le crayon lors de l'étude de documents.

iBase : Gestionnaire de bases de données d'analyse. Entièrement intégré avec les outils d'analyse : Analyst's Notebook, DataMiner, PatternTracer, SIG. Sert à la fois à stocker les

informations, à les mettre à disposition des utilisateurs non informaticiens, et à gérer le transfert automatique entre ces données et les outils d'analyse quelqu'ils soient. iBase permet de manipuler les données en dessinant sous forme d'objets et les liens les hypothèses de travail de l'utilisateur final.

DataMiner : outil statistique intégré verticalement avec iBase.

PatternTracer : Recherche automatisé de comportements répétitifs entre des transactions téléphoniques ou bancaires.

Interface SIG : Permet la liaison bi directionnelle entre iBase et les principaux Systèmes d'Information Géographique (Mapinfo, SerView, BLUE8).

INTELLIXIR SARL www.intellixir.com	• veille Internet • **recherche et indexation** • **text mining** • **data mining** • traduction • traitement de l'image • représentation graphique • knowledge management	

Adresse :
2 Boulevard de la Plaine, 04100 Manosque

Date de création : 2002

Partenariats :
CEA

Produits :
Intellixir : logiciel spécialisé dans la bibliométrie et l'infométrie décisionnelle. Il a été développé au sein d'un équipe du CEA de CADARACHE (plate-forme logicielle SIMBAD) et reste la propriété du CEA.
Il analyse de manière statistique les données documentaires dans le domaine scientifique en permettant d'établir des corrélations entre chercheurs, auteurs de publications et entreprises dépositaires de brevets.
La source documentaire provient d'entreprises spécialisées dans le recueil de publications.

Abonnement annuel : 20.000 €

ISCOPE www.iscope.fr	• **veille Internet** • **recherche et indexation** • **text mining** • **data mining** • traduction • traitement de l'image • **représentation graphique** • knowledge management	

Adresse :
19 - 21, rue Valette – 75005 PARIS
Siège social : 9, avenue Léon Marchand – 94 320 THIAIS

Date de création : 2000

Partenariats :
Atlantic Intelligence, TEMIS

Concurrents :
BEA Conseil

Produits :
Keywatch : Crawl et surveillance : collecte de l'information la plus large possible afin d'assurer une couverture totale et de pallier le problème de l'information volatile.
La collecte est effectuée sur les sources Internet veillées et éventuellement sur les liens pointés par les sources (profondeur de crawl paramétrable).
L'actualisation de l'information est assurée selon une fréquence allant de la minute au mois selon le type de source. Le différentiel entre 2 aspirations est identifié finement (capacité à ne pas prendre en compte des changements d'annonce publicitaire, par exemple).
La solution répond à la totalité des difficultés rencontrées en matière de « crawl » : web invisible (gestion de l'accès à des services commerciaux), codes mobiles, java, flash, mail, pièces jointes attachées… sans développement spécifique, le paramétrage des sources étant générique.
Elle peut être utilisée en mode ASP ou sur un serveur du client.
Le filtrage et l'indexation fonctionnent sur requêtes booléennes, qui peuvent être complétées par une fonction de stemming[1] et la définition par l'utilisateur d'équivalence de mots ou d'expressions. Une fonction d'alerte est assurée selon les profils déterminés par les utilisateurs. Le système assure le dédoublonnage.

La plate-forme KeyWatch intègre également des fonctions d'analyse de l'information :
- une fonction de data-mining relativement simple associée à un module de visualisation,
- sur demande du client, Keywatch peut également intégrer des outils de text-mining d'autres éditeurs, notamment ceux de TEMIS ou de SPSS.

[1] Stemming : lemmatisation générique, soit l'estimation automatique de la racine des mots, sans dictionnaire. Cette méthode est bien plus rapide et moins coûteuse que l'analyse morphologique basée sur des ressources linguistiques, mais naturellement moins précise. Les langues actuellement supportées par la fonction stemming de Keywatch sont les suivantes : anglais, français, italien, espagnol, néerlandais, finnois, norvégien, suédois, danois, portugais, japonais (en cours).

KARTOO www.kartoo.com www.kartoo.net www.ujiko.com	• **veille Internet** • **recherche et indexation** • text mining • data mining • traduction • traitement de l'image • **représentation graphique** • knowledge management	

Adresse :
10 allée Evariste GALLOIS, Parc Technologique de la Pardieu à Clermont-Ferrand (63).

Date de création : 10/10/2001

Partenariats :

Concurrents :
le groupe américain ESRI.

Produits :
KartOO Veille : Surveillance de sites, veille sur moteurs, gestion des favoris, base de connaissance commune, cartographie

KARTOO : méta moteur de recherche sur Internet. Particularité : propose non pas une liste de sites mais une carte avec les principaux liens qui mènerons les internautes vers la réponse qu'ils cherchent à obtenir.
Sur cette carte, KARTOO positionne les sites du centre vers la périphérie en fonction de leur proximité avec l'objet de la recherche. Au final, l'utilisateur dispose d'une sorte d'arborescence dans laquelle il est possible de naviguer en cliquant sur chaque thème.

UJIKO : nouvel outil (2004) de recherche basé sur un index de 4 milliards de documents. Ce nouveau concept offre des fonctionnalités de mémorisation et de filtrage permettant aux utilisateurs de le personnaliser facilement. Les résultats obtenus avec ce nouveau produit ont été jugés au moins aussi pertinents que ceux de Google par les spécialistes des moteurs de recherche.

Kartoo Site Box : Représentation graphique de résultat de recherches sur moteur de recherche d'un site Web
Kartoo Intranet : Représentation graphique de résultat de recherches sur moteur de recherche interne
Kartoo BDD : Représentation graphique de recherches dans des bases de données
Kartoo GED : Méta-moteur de recherche interne sur sources de données hétérogènes et représentation graphique associée
Kartoo Visualisation : Seule réelle innovation : Représentation graphique de données structurées au format XML (concepts etc.) produit principal de la société Kartoo sur lequel s'appuient tous les autres produits.
Kartoo KM : Offre de service de gestion de la connaissance.

Atouts technologiques :
Personnalisation par mémorisation du profil utilisateur

KNOWINGS www.knowings.com	• **veille Internet** • **recherche et indexation** • text mining • data mining • traduction • traitement de l'image • représentation graphique • **knowledge management**	

Adresse :
BP 354 – Savoie Technolac – 73372 Le Bourget du Lac Cedex

Date de création : 29.04.1999

Partenariats :
ANELIA, CEGOS, Agence virtuelle Suisse

Concurrents **:**

Produits :
Knowledge manager : Management de la connaissance et e-collaboration
Global finder : Crawl, classement, diffusion

KOPPERTI www.KopperTi.com	• veille Internet • **recherche et indexation** • text mining • data mining • traduction • traitement de l'image • **représentation graphique** • knowledge management	

Adresse :
31, avenue du Parc – 94340 Joinville le Pont

Date de création : 30 janvier 2004

Partenariats :

Concurrents :
Géoconcept (Fr), Mapinfo (US), Srim

Produit et service :

Solution permettant la réalisation de tableaux de bord avec une représentation graphique sur fonds de cartes géographiques. La société s'inscrit dans la catégorie des SIG : systèmes d'information géographique).

L'outil, destiné aux directions marketing, commerciales et stratégiques, permet d'interroger des bases de données sur des clients, fournisseurs et de voir les résultats de l'analyse sur un fonds de carte (France, Europe, Monde).

La solution semble plutôt bien construite et simple d'utilisation. Elle a également la particularité de fonctionner en continu et donc de pouvoir être interrogée à tout moment ; contrairement à certains outils ne permettant que la réalisation de points de situation ponctuels.

Elle fonctionne soit en mode service (ASP), la société remet des rapports réguliers au client, soit en mode fournisseur produit : le produit est vendu puis simplement paramétré par un consultant.

LINGWAY www.lingway.com	• veille Internet • **recherche et indexation** • **text mining** • data mining • **traduction** • traitement de l'image • représentation graphique • knowledge management	

Adresse :
33-35 rue Ledru Rollin Bât C, 94 Ivry / Seine

Date de création : 2001

Partenariats :
JOUVE, KEO, Oriam, Parisbiotech,…

Concurrents :
Sinequa, Temis, Convera, Triplehop,…

Produits :
LINGWAY commercialise un "package" de trois produits qui peuvent aussi être vendus séparément : un moteur d'analyse sémantique, un moteur d'indexation et un dictionnaire électronique (5 langues).
Jusqu'à présent, le gros du Chiffre d'affaires de la société était réalisé avec des organismes publics (INPI, BNF,..). Avec son offre packagée, LINGWAY entend s'attaquer au privé.

Le point fort du produit est la présence de concepts (150 000) associés à des dictionnaires de langues (5 actuellement). La finalité est de pouvoir faire des recherches sur un mot ou une expression est d'avoir le résultat dans les 5 langues avec les textes correspondants (articles de presse, Internet,...).

7 ou 8 des 16 employés ont une expérience du text mining d'au moins 15 ans.

LTU TECHNOLOGIES www.LTUtech.fr	• veille Internet • **recherche et indexation** • text mining • data mining • traduction • **traitement de l'image** • représentation graphique • knowledge management	

Adresse :
57, rue Pierre-Charron 75008 Paris
Filiale à 100% à Washington – USA

Date de création : 1999 par des chercheurs issus du MIT Media Lab., de l'Université d'Oxford et de l'INRIA

Partenariats :
DAVINCI Consulting Tecnologico – Scube – Global Linxs – MelonWeb – TKC Communications – Jouve – EDS

Concurrents :
Convera (USA), Clearswift (filtrage d'images)

France Télécom, Sagem, Adamentium (protection des mineurs)

Produits :
Image-seeker : système complet de gestion des images avec indexation et recherche par le contenu.
Les fonctionnalités :

- Recherche par le contenu : texte et image combiné, navigation par similarité, soumission d'images nouvelles, recherche avancée couleur/forme, signature ADN spécifiques pour certains domaines d'applications, outils de sélection de requêtes sur une partie de l'image, analyse automatisée de grands volumes de fichiers, génération de rapports, réponses formatées en XML
- Gestion des préférences utilisateurs, historique des requêtes, sauvegarde images et recherches
- Administration de contenu : interface web pour l'indexation aisée de répertoires d'images, intégralement configurable, indexation de vidéos, export de données
- Administration de système : droits utilisateurs, transmission sécurisée des images, surveillance de l'application

Image-filter : logiciel de gestion de contenu visuel permettant l'analyse automatique des images et le filtrage de contenu à caractère litigieux.
Les fonctionnalités :

- Indexation d'images en temps réel
- Reconnaissance d'images à la volées
- Classification et renvoi du résultat
- Définition de « blacklists / whitelists »
- Niveau de tolérance modulable

Image Indexer : indexation.

LTU propose des adaptations de sa plate-forme d'analyse d'images pour répondre aux besoins spécifiques des clients. Dans ce cadre, LTU propose des produits aux fonctionnalités « avancées » :

- suggestion automatique de mots-clefs :
  - automatisation du processus de classification d'images
  - standardisation des mots-clefs associés aux nouvelles images, en s'appuyant sur les mots-clefs présents dans une archive d'images annotées
  - réduction des coûts engendrés par une annotation manuelle image
- automatisation de la vidéo
- applications audiovisuelles : détection temps-réel d'images clefs prédéfinies dans un flux vidéo
- « clipping, archiving and repurposing applications » : segmentation temporelles de la vidéo pour l'archivage et la publication
- pige vidéo : pige quantitative et qualitative et contrôle des flux vidéos protégées par un droit d'auteur.

MANREO www.manreo.com	• veille Internet • **recherche et indexation** • text mining • data mining • traduction • **traitement de l'image** • représentation graphique • knowledge management	

Adresse :
6, boulevard Saint Denis – 75010 PARIS

Date de création : 1999

Partenariats:
Commerciaux : Sony – DPS France – TV-RADIO.COM – VIEWON TV – AAVS – HYBRID MC – HYBRID MC – PENNA MEDIA LTD
Techniques : Real Network – Microsoft Windows media Player – Macromedia

Concurrents :
KINOMAI et LTU.

Produits
Manréo propose 6 packages :

- Pour créer :
  - Manréo CAFÉ ayant pour but d'automatiser complètement la capture et l'encodage de contenus vidéos grâce à la détection, dans un flux vidéo, d'images clefs préalablement renseignées,
  - Hypercast Editor pour décrire, annoter, indexer de façon globale et temporelle et cataloguer n'importe quel contenu audio-vidéo,
  - Hypercast Live pour produire des présentations par production et publication de documents Web, synchronisés sur de la vidéo.
- Pour gérer : Hypercast Warehouse pour centraliser et partager les archives audio et vidéo, alimenter un portail Web et d'optimiser le processus de montage.
- Pour distribuer : Hypercast Publishing Server pour automatiser et fiabiliser la diffusion des contenus multimédias.
- Pour le design : Hypercast TempleMaker pour créer des modèles de présentations Rich Media, modèles compatibles avec la majorité des navigateurs du marché.

MAPSTAN

Nom légal : VOYEZ VOUS
www.mapstan.net et www.mapstan.com

- **veille Internet**
- **recherche et indexation**
- **text mining**
- data mining
- traduction
- traitement de l'image
- **représentation graphique**
- knowledge management

Adresse :
VOYEZ-VOUS SA, 52 Boulevard Sebastopol - 75003 PARIS 03

Date de création : 2000

Partenariats :
Colt pour la bande passante, mais Mapstan se charge de l'hébergement.

i-KM, cabinet de conseil spécialisé dans la gestion de la connaissance.

Concurrents :
Mapstan se positionne sur le même segment que les outils Kartoo (cartographie de l'information en interface flash).

Produits :
A l'issue d'un programme de R&D soutenu par l'ANVAR, MapStan a développé une technologie "100% Java" de cartographie contextuelle, le Web Positioning System™ (WPS) qui est au cœur de ses outils.

Pour valider sa technologie à grande échelle, MapStan a réalisé deux démonstrateurs mis à la disposition de tous les internautes. Lancé en septembre 2001, MapStan.net est le premier service de navigation personnalisée sur Internet. Depuis janvier 2002, MapStan Search fournit une synthèse visuelle des résultats et recommandations issus de recherches sur Internet. MapStan Ces services ont été largement plébiscités par la presse informatique.

Mapstan propose une gamme de solutions pour collecter (MapStan Search), accéder (MapStan Intranet), partager (Mapstan Groups) et analyser (Mapstan Analytics) l'information à l'aide de technologies cartographiques.

Une version ASP pour sa solution MapStan Search s'interface avec des moteurs de recherche, sur Internet ou Intranet, et permet de mutualiser et de capitaliser les recherches effectuées par un groupe d'utilisateurs. Concrètement, MapStan Search propose la représentation des résultats d'une requête sous la forme d'un plan de quartier sur lequel les rues, plus ou moins larges, symbolisent la "proximité" des réponses reçues et donc leur pertinence par rapport à un profil d'utilisateur déterminé. Ainsi, pour une recherche donnée, Mapstan Search, qui archive systématiquement toutes les requêtes et leurs résultats, présente les résultats de recherches similaires déjà effectuées.

La version Entreprise de Mapstan Search est commercialisée à partir de 15 000 euros par processeur.

MATHEO SOFTWARE www.matheo-software.com	• **veille Internet** • recherche et indexation • text mining • **data mining** • traduction • traitement de l'image • **représentation graphique** • knowledge management	

Adresse :
69, rue du Rouet, 13008 Marseille

Date de création : 29/01/2003

Partenariats :
Distributeur : SARL IMCS (Conseil pour les affaires et la gestion – Gérant : M. Jean-Marie DOU).

Concurrents **:**

Produits :
MATHEO SOFTWARE développe 3 logiciels dans les domaines de compétences de la veille, de la bibliométrie, de la cartographie d'informations et des nouvelles technologies :

MATHEO ANALYZER : logiciel de traitement automatisé de l'information: importation de données, traitements statistiques, visualisations graphiques, création de réseaux, matrices. Ce produit est loué 600 euros par an,
MATHEO PATENT : logiciel de veille et de traitement de l'information 'brevets'. Ce produit est vendu 3.500 euros.
MATHEO WATCH : produit d'appel non vendu mais proposé à des chercheurs dans le domaine médicale. Logiciel de veille sur site, il permet de surveiller des URL présélectionnées, et doit être régulièrement exécuté pour identifier toutes les modifications apportées sur les sites pré-définis.

MONDECA www.mondeca.com	• veille Internet • **recherche et indexation** • **text mining** • data mining • traduction • traitement de l'image • représentation graphique • **knowledge management**	

Adresse :
3, Cité Nollez 75018 PARIS

Date de création : Juin 99

Partenariats :
Unilog, Euriware, Sonovision Itep, PCO Technologies, Erin,Temis, IBM, BEA, Oracle

Concurrents :
Américains : Semagix, Unicorn, Stratify

Produits :
Mondeca est un éditeur de logiciel spécialisé dans les solutions de gestion des connaissances et organisation des ressources documentaires.

La solution qu'elle développe se nomme ITM (Intelligent Topic Manager). ITM est une solution logicielle de management de l'information. Basé sur une modélisation par ontologie (organisation hiérarchique de la connaissance sur un ensemble d'objets par leur regroupement en sous-catégories), ITM propose un accès à des contenus d'informations au travers d'un portail web sémantique. Par ailleurs, ITM offre la gestion de tous les éléments d'organisation et de structuration des ressources d'informations. Mondeca cible tous les comptes pour lesquels le fonds documentaire est un levier fondamental : pharmacie, finance, administration, édition, sociétés de service et de conseil.

NEWPHENIX
www.new-phenix.com

- veille Internet
- **recherche et indexation**
- **text mining**
- data mining
- traduction
- **traitement de l'image**
- traitement de la parole
- représentation graphique
- knowledge management

Adresse :
33 Rue galilée 75016 Paris

Date de création : 2003

Partenariats :
Oryx, CEA

Concurrents :

Produits :

Phenix Multimédia Multilingual Engine : moteur d'analyse et de modélisation de l'information multimédia (texte et image).

Phenix Enterprise Search : serveur de recherche utilisant langage naturel et crosslingues.

Phenix Profiling : serveur de filtrage de flux d'information multimédia, diffusion sélective d'information, applications de surveillance de la confidentialité de l'info.

Secteurs ciblés : l'intelligence économique, la gestion de la relation client, l'e-business, la gestion de projet, la recherche et développement, les portails d'entreprise, la gestion de contenus.

NOEMATICS www.noematics.com	• veille Internet • **recherche et indexation** • **text mining** • data mining • traduction • traitement de l'image • représentation graphique • knowledge management	

Adresse :
1 rue Albert EINSTEIN 77420 Champs sur Marne.

Date de création : 1995 (sous le nom d'Ermen)

Partenariats :
Diatopie, Mixio, Elitbureau, Imme

Concurrents :

Produits :
Reflexion : Moteur d'indexation français-anglais

NOHETO www.noheto.net	• **veille Internet** • **recherche et indexation** • **text mining** • data mining • traduction • traitement de l'image • traitement de la parole • représentation graphique • knowledge management	

Adresse :
101/103 boulevard Mac Donald 75019 PARIS

Date de création : 2000

Partenariats :
Atos Origin, Cap gemini, IBM, Sopra, Bea, IBM, Verity

Produits :

WCM Manage Server : moteur de recherche, indexation, classification, hiérarchisation. diffusion interne et externe.

Gestion de contenu.

ORBISCOPE www.orbiscope.net/fr	• **veille Internet** • **recherche et indexation** • text mining • data mining • traduction • traitement de l'image • représentation graphique • knowledge management	

Adresse :

Date de création :

Partenariats :

Concurrents **:**

Produits :
Orbiscope meta recherche : Crawl et surveillance
Orbiscope Position : Surveillance de positionnement dans les engins de recherche

ORDIMEGA www.ordimega.com	• veille Internet • **recherche et indexation** • text mining • **data mining** • traduction • traitement de l'image • **représentation graphique** • knowledge management	

Adresse :
87 bis, avenue de Wagram - 75017 Paris

Date de création : 1986

Partenariats :
Diane-ORT, Coface

Concurrents :
Sage, Cegid

Produits :
- *Préface* (historique, prévision, évaluation expert, EAO) : 7 555 € pour le module complet en licence monoposte. Les 5 modules peuvent être achetés séparément. Ils nécessitent, par ailleurs l'acquisition de bases de données commerciales (Diane, Coface-ORT).
- *Préface OSB* : logiciel de benchmarking financier proposé par abonnement annuel monoposte (3 450 € HT)

Préface permet l'analyse de plusieurs dizaines d'agrégats, scoring et ratios financiers d'entreprises, de les commenter en se basant sur des milliers de cas étudiés et de les positionner par rapport à leur environnement (sectoriel, géographique, taille, etc). Le logiciel fournit de nombreuses représentations graphiques et indique les forces et faiblesses des entreprises.
Il permet également de faire des prévisions sur l'évolution d'une entreprise et d'apprécier sa valeur réelle. La construction de bilans fictifs simulera des évolutions et aidera à la prise de décisions.
Les produits proposés sont hétérogènes, tant sur le plan de la forme que sur celui des objectifs. Ils permettent de juger, d'expliquer, de suggérer des décisions mais ils jouent aussi un rôle formateur, donnant la formule de calcul et la définition de chaque indicateur, ainsi que ses relations avec d'autres éléments. En interrogeant ces derniers, l'utilisateur approfondit sa perception de la situation de l'entreprise.

PERTIMM www.pertimm.fr	• **veille Internet** • **recherche et indexation** • **text mining** • data mining • traduction • traitement de l'image • représentation graphique • knowledge management	

Adresse :
20, rue Montesquieu 92600 Asnières sur Seine

Date de création : 1997.

Partenariats :
ALOGIC, DALTATECH, PERTINENCE MINING, Newpartner international, Ecole des mines, Thales,…

Distributeur : ENNOV, CREATEAM, SOFTLAB, DELTATECH, QUESTEL, ORBIT, ALOGIC, OVERLAP, OFYE, NEWPARTNER, PERTINENCE, TECH SOURCE, TEXTEC

Concurrents :
Verity (US), Autonomy (GB), Sinomia, Sinéqua, Spirit,…

Produits :
Pertimm : GED et moteur d'indexation et de recherche sémantique, plein texte, classification, langage naturel, logique booléenne, logique floue, expressions, cross-lingue et PERTIMMIZERS (systèmes de pointeurs d'un ou plusieurs utilisateurs sur un thème donné).

Pertimm Smart Drive : Pertimm sur un réseau local pour permettre aux utilisateurs de retrouver et exploiter leurs données communes et/ou personnelles.

Pertimm Web : Pertimm sur Intranet permettant de gérer plusieurs millions de documents.

Pertimm Notes : permettant aux utilisateurs de Lotus Notes d'accéder à toutes les bases en même temps mais aussi à des serveurs http classiques.

Projet en cours: Grabbing (aspiration de sites) et différentiel (identification de l'ensemble des nouvelles pages sur un site).

Facturation : Location annuelle ou vente en fonction du nombre d'utilisateurs et du nombre de fichiers a indexer.

Plusieurs brevets portant sur les « signatures », les « référents » (sous-ensemble contenants de l'information et leur moyen d'accès), les « vecteurs topographiques » (position de la signature comme référent unique) et les « cross-références » permettant de faire des recherches croisées sur l'ensemble des contenus multimédia.

PERTINENCE MINING
www.pertinence.net

- **veille Internet**
- recherche et indexation
- **text mining**
- data mining
- traduction
- traitement de l'image
- représentation graphique
- knowledge management

Adresse
82, rue Jean Jaurès 94400 Vitry sur Seine

Date de création : 22.05.2002

Partenariats :
Technologiques : Adobe, Pertimm, Exalead

De contenu : NewsPresse, PrNewswire, AFP et syndicat de presse des quotidiens régionaux

Concurrents :

Produits :
Pertinence summarizer (PS) : logiciel de résumé automatique.

Pertinence Information Network (PIN) : plate-forme de collecte, de traitement et de diffusion.

KENiA (Knowledge Extraction and Notification Architecture) est une marque et une technologie propriétaire de Pertinence Mining. Ce logiciel permet la surveillance des contenus des sources d'information.

Pertinence News Extractor (PNE) : module propriétaire permet l'alimentation de la plate-forme PIN. L'extraction automatique est opérationnelle en 14 langues.

PULSAR NOTRINO www.pulsarnotrino.com	• **veille internet** • recherche et indexation • text mining • data mining • traduction • traitement de l'image • **représentation graphique** • **knowledge management**	

Adresse :
11 rue du Commerce, 75015 PARIS

Date de création : 2004

Partenariats :
MAPSTAN, AMOWEBA

Produits :
BENCHMARK Marketing est une solution de veille dans le domaine du marketing concurrentiel. Le produit est actuellement en test dans les grandes agences de communication françaises. Il compare l'image et le positionnement d'une marque dans son secteur et évalue la performance d'un client vis à vis de ses concurrents directs.

QWAM www.qwam.com	• **veille internet** • **recherche et indexation** • text mining • **data mining** • traduction • traitement de l'image • traitement de la parole • représentation graphique • knowledge management	

Adresse :
40, rue des vignobles, 78400 Chatou

Date de création : 1997

Partenariats :
Pas de partenaires attitrés, mais de relations contractuelles liées en fonction de réponses à des marchés spécifiques.

Produits :
Solutions logicielles de fédération, d'intégration et de gestion de contenus électroniques. QUAM E-CONTENT SERVER (QES) : modules fonctionnels pour les applications informationnelles. Il s'agit d'une plate-forme logicielle d'accès, recherche, consultation, diffusion et gestion des contenus électroniques pour intranets/extranets.

SEMIOSYS http://www.semiophore.net	• veille Internet • recherche et indexation • **text mining** • **data mining** • traduction • traitement de l'image • **représentation graphique** • knowledge management	

Adresse :
7 rue des roses, 85100 Les Sables d'Olonne

Date de création : 27/10/2000

Partenariats :

Concurrents **:**

Activité :
Initialement : solutions d'analyse automatique de contenu puis s'est orienté vers la visualisation.

Produits :
Traitement du langage :
Semioxtractor : extraction et filtrage de syntagmes nominaux, résumés de texte paramétrables.
Semioclust : auto organisation thématique, éventuellement hiérarchisée, de collections de documents ou d'objets textuels.

Analyse et visualisation (éclairage et mise en perspective de l'information) :
Semiophore Explorer : Plateforme de visualisation générique orientée réseaux (données explicitement relationnelles).
Semiophore® Miner : Semiophore Miner est un logiciel d'analyse et de visualisation de documents, préférablement structurés (XML ou base de données) qui permet de naviguer dans les documents à travers des réseaux de collaborations ou des réseaux de concepts qui sont générés par le logiciel. Il est principalement utilisé sur des bases de données bibliographiques et bases de brevets mais peut traiter tout type de document. La partie
>analyse de données fait appel à des techniques TAL d'extraction et d'analyse sémantique. La partie navigation repose sur des techniques de visualisation interactive et de mise en perspective des résultats selon différents points de vue, notamment des mesures d'analyses de réseaux sociaux et la prise en compte de la dimension chronologique (animations longitudinales par exemple).

Semiophore Multimaps : Outils de catégorisation/hiérarchisation/organisation manuelle et cartographie de l'information.

SINEQUA www.sinequa.com	• veille Internet • **recherche et indexation** • **text mining** • data mining • traduction • traitement de l'image • représentation graphique • knowledge management	

Adresse :
51-59, rue Ledru-Rollin, 94200 Ivry-sur –Seine

Date de création : 14/03/1984

Partenariats :
Thalès (intégrateur de SINEQUA), Net2One

Concurrents:
Exalead, Pertim, Arisem, Verity, Convera, Autonomy

Produits :
La gamme de produits repose sur un moteur d'indexation nommé Intuition
Intuition effectue une indexation linguistique dans 15 langues :
Français, Anglais, Allemand, Espagnol, Italien, Néerlandais, Japonais, Chinois traditionnel, Thaïlandais, Danois, Finnois, Grec, Coréen, Portugais, Russe, Suédois

Intuition : moteur d'indexation et de recherche. Mis au point en 98, amorcé en 94 avec un projet Eureka
Il se distingue des autres moteurs par l'emploi de techniques linguistiques avancées:
- moteur d'analyse morpho-syntaxique
- algorithme sémantique innovant mettant en œuvre un dictionnaire sémantique très volumineux et un mode de stockage vectoriel
- une vaste gamme de fonctionnalités de recherches avancées
- la mise en œuvre de fonctions de conceptualisation et de reconnaissance d'entités qui permettent de naviguer intelligemment dans l'information.

Le moteur se veut polyvalent grace à l'utilisation de plusieurs type de recherche : utilisant la linguistique (lexicologie et sémantique), les statistiques, la mathématique et la recherche . Différentes fonctions de navigation permettent de rebondir sur une liste de réponses, afin de recentrer ou de réorienter le sujet d'une requête. Recherche par l'exemple, affinage, navigation par concepts ou par entités.

Suite logicielle :
• **iIntranet** : solution intranet
• **iInternet** : indexation de sites et de bases de données
• **iCatalog :** mise en place de catalogues en ligne
• **iPush :** diffusion sémantique et sélective de l'information
• **iCédérom :** moteur pour Cédérom

Enfin, pour permettre aux utilisateurs d'embarquer un moteur sémantique dans leurs applications, Sinequa a développé une véritable boîte à outil sémantique : **iToolbox**.

SOFTISSIMO www.softissimo.com www.reverso.com	• **veille Internet** • recherche et indexation • **text mining** • data mining • **traduction** • traitement de l'image • représentation graphique • knowledge management	

Adresse :
33, av Mozart – 75016 Paris

Date de création : 1987

Partenariats :
Verity pour les moteurs de recherche
ProjectMT (logiciel de traduction)

Concurrents :
Systran

Produits :
Reverso : logiciel de traduction automatique en français, anglais, allemand, espagnol
LEXIBASE Collins : dictionnaire électronique.

Les dictionnaires sont spécialisés par industrie : automobile, pétrole, médical, informatique, finance…
Les logiciels de traduction sont proposés dans les différentes langues du Moyen Orient, de l'Asie et européennes. La solution multilingue permet l'écriture avec les caractères de plus de cent langues (chinois, arabe, russe, etc…) directement dans Word.

Le dictionnaire personnel peut être enrichi : possibilité d'ajouter des mots ou expressions supplémentaires. Les expressions vont du langage soutenu à l'argot. Par la recherche approximative, le logiciel comprend en général ce que l'utilisateur recherche. La navigation est simple et le passage d'une langue à l'autre est très facile. Le plus, réside dans la fonction automatique pour passer directement d'un fichier de traitement de texte au dictionnaire ;
En matière de veille technologique, le logiciel Reverso permet rapidement le décodage d'un texte ou la traduction à la volée de pages Web.

SYNOMIA www.synomia.fr	• veille Internet • **recherche et indexation** • text mining • data mining • traduction • traitement de l'image • représentation graphique • knowledge management	

Adresse :
3 rue nationale – 92 100 Boulogne Billancourt

Date de création : 10 août 2000.

Partenariats :
Partenariat exclusif de développement technique et d'exploitation commerciale avec le CNRS.
Partenaire d'IBM pour ses serveurs informatiques et de l'opérateur international TELIA pour leur hébergement.
ANVAR

Concurrents :
produits Gold Search de Google, Atomz, HTDIG…

Produits :
Synomia Search : moteur de recherche et d'indexation de site web qui regroupe automatiquement les résultats en fonction de leur sens et en identifiant les sous-catégories apparues par une expression représentative de leur contenu. Synomia analyse les résultats en temps réel et assure leur classification permettant la couverture de l'intégralité du contenu d'un site internet ou d'une seule rubrique. Dans son analyse, le moteur se réfère à la structure éditoriale existante dont il reprend le contenu document par document. Synomia utilise une analyse morpho-syntaxique pour s'affranchir des variations linguistiques non significatives. Une aide à la recherche est fournie à l'utilisateur sous la forme de vues d'ensemble des différentes formulations de l'information.

Facturation : l'abonnement mensuel varie de 300 et 1200 euros.

SYSTRAN
www.systransoft.com

- **veille Internet**
- recherche et indexation
- text mining
- data mining
- **traduction**
- traitement de l'image
- représentation graphique
- knowledge management

Adresse :
Paris Nord - La grande arche
1, Parvis de la Défense
92044 PARIS LA DEFENSE CEDEX
Date de création : 30/01/1986

Partenariats :

Concurrents :
PRO MT (éditeur russe distribuant en France par l'intermédiaire de SOFTISSIMO sous le nom de REVERSO) et ALIS TECHNOLOGY (société canadienne)

Produits :
Outils de traduction automatique facilitant la traduction dans 52 paires de langues, dont 12 comprenant le français et dans 20 domaines spécialisés.
La technologie est basée sur des dictionnaires (dictionnaires généraux + 20 dictionnaires thématiques, l'utilisateur pouvant également créer des dictionnaires personnels) et des règles linguistiques.

TECHNOLOGIES - GID
www.t-gid.com

- veille Internet
- **recherche et indexation**
- text mining
- data mining
- traduction
- traitement de l'image
- représentation graphique
- knowledge management

Adresse
84, boulevard de la Mission Marchand – 92411 COURBEVOIE

Date de création : 1984

Partenariats :
Aucun pour le produit « Spirit »

Concurrents :
Toutes les sociétés de GED

Produits :
Rsc2000 : surveillance multi plates-formes à distance des systèmes, applications, réseaux et services

Conceptor : génère une documentation à partir du système d'information

Spirit : effectue une indexation automatique en texte, créant ainsi des bases informationnelles à partir de documents internes et sources différentes (sites Internet, fichiers bureautiques, bases GED). Un dictionnaire est inclus permettant la traduction automatique du français, de l'anglais, du néerlandais. Ce produit ne nécessite pas une formation lourde, est paramétrable selon les vœux du client.

TEMIS
www.temis-group.com

- veille Internet
- **recherche et indexation**
- **text mining**
- data mining
- traduction
- traitement de l'image
- représentation graphique
- knowledge management

Adresse :
Tour GAMMA B 193-197 Rue de Bercy 75582 Paris
Présent en Allemagne et en Italie

Date de création : 2000

Partenariats :
Iscope, Modeca
IBM, SWORD, AKAMAI,

Intégrateurs français : Atos Origin, Ifatec, Overlap, Unilog.

Sociétés incluant la technologie Temis dans leurs applications logicielles : Iscope, Koltech , Mondeca, Xylème, Ontoprise, Lightobjects, AskMe, IBM Life Sciences
Universités françaises (laboratoires et DESS) et étrangères.

Concurrents :
Inxight (américain)
Clearforest (israelien)
SPSS, Lingway

Activité :
TEMIS (TExt MIning Solutions) propose des outils de Text Mining.

Produits :
Insight Discoverer Extractor : serveur d'extraction d'information dédié à l'analyse de documents textuels non structurés. Il détecte les concepts qui ont été définis pertinents pour les utilisateurs comme par exemple une annonce de fusion dans un article pour un analyste, une opportunité commerciale dans un e-mail pour un chargé de clientèle, une compétence spécifique dans un CV pour un recruteur. Il lit les documents électroniques dans plus de cinquante formats différents.

Insight Discoverer Categorizer : serveur de catégorisation de documents. Il classe automatiquement les documents non structurés dans des catégories prédéfinies, en combinant des règles d'analyse statistiques et linguistiques.

Insight Discoverer Clusterer : serveur de classification automatisée qui regroupe dynamiquement les documents en fonction de leur ressemblance sémantique et de leur proximité thématique. Il propose le classement le plus pertinent pour un fonds documentaire donné. Les utilisateurs peuvent ainsi naviguer dans leurs documents organisés par thèmes et sous-thèmes.

XeLDA : moteur linguistique multilingue. Il modélise et normalise des documents non structurés, en vue d'une exploitation automatique de leur contenu. 18 langues européennes sont actuellement traitées.

XTS : suite logicielle qui propose une assistance dans la création de terminologies d'entreprise et les exploite pour améliorer la cohérence et la qualité de productions documentaires.

THALES Land & Joint Systems www.thales-communications.com	• **veille Internet** • **recherche et indexation** • **text mining** • **data mining** • traduction • **traitement de l'image** • **représentation graphique** • **knowledge management**	

Adresse :
160 boulevard de Valmy, BP 82, 92 704 Colombes

Partenariats :
Sinequa, Arisem

Concurrents :
EADS, France Telecom, Autnonomy

Produits :
Kaliwatch Professional : Collecte d'information multi-sources (web, e-mail, forums, bases documentaires, etc.), catégorisation, extraction. Multi et cross linguisme. Mono-utilisateur. Paramétrage métier (thésaurus).

Kaliwatch Entreprise : Portail multi-utilisateurs. Outre les fonctionnalités de la version « Pro », la version « Entreprise » propose également des fonctionnalités de gestion de base de connaissance commune, de diffusion « push », des fonctions de gestion des droits d'accès, et certaines fonctionnalités de mining plus avancées. Cette version est conçue pour traiter des gros volumes d'informations, en provenances de sources hétérogènes et diversifiées.

La technologie Kaliwatch est linguistique (dictionnaires, analyse morpho-syntaxique, lemmatisation, index de désambiguïsation). La plate-forme utilise les standards J2EE (Java, EJB, JSP, servlet, etc.) et SQL.

Idéliance : Gestion de « réseaux sémantiques » (réseaux de phrases simples dont les nœuds sont constitués par les groupes de mots communs). Module de visualisation.

IRIS : la société Thales Communication entend, au travers du projet IRIS, se positionner comme intégrateur national d'outils de traitement de l'information, notamment pour les services de renseignement. Le projet consiste à développer un socle technique commun, porteur d'un ensemble de modules de traitement documentaire avancé (catégorisation, extraction d'entités nommées et surlignage, résumé automatique, attribution d'indices de gravité), de visualisation graphique, dans l'espace (localisation des évènements sur carte) et le temps (visualisation chronologique) et d'aide à l'analyse (classification, statistiques, extraction de réseaux). A terme, des capacités de traitement de la parole et de l'image sont prévues.
Thales Communications annonce vouloir intégrer, après sélection, les produits les plus performants de l'offre nationale dans une démarche partenariale.

TRIVIUM www.trivium.fr	• **veille Internet** • recherche et indexation • **text mining** • data mining • traduction • traitement de l'image • **représentation graphique** • **knowledge management**	

Adresse :
12, boulevard de Sébastopol – 75004 PARIS

Date de création : 1992

Partenariats:
Accenture, Cap Gemini, IBM, Ernst & Young, Euriware

Concurrents:
People Soft, A.I.M..(U.S)

Activité :
TRIVIUM est un éditeur de logiciels ayant pour la spécialité la cartographie d'informations et dont la politique de marketing vise plus particulièrement la gestion des ressources et les compétences humaines des entreprises. Les outils de Trivium peuvent cependant s'adapter à la veille.

Produits :
SEE-K, solution permettant la structuration d'informations non organisées et les restituant sous forme graphique : concept de « l'arbre de compétences ».
Cette solution intègre les produits Gingo et Umap, qui n'existent plus en tant que tels.

UNILOG www.unilog.com	• veille Internet • recherche et indexation • text mining • data mining • traduction • traitement de l'image • représentation graphique • knowledge management	

Adresse :
37 rue du Rocher – 75738 PARIS Cedex 8

Date de création :
1968 (sous le nom Informatique et entreprise), 1983 (création d'Unilog)

Partenariats :
Alliances avec Keane et Escan.
Partenaires offre KM :
KM : Knowesis, Sx-Sigma
Editeurs outils de veille : Arisem, Exalead, Temis, Mondeca, Inxight, Autonomy, Verity, Agilience, Entopia, Sinequa
Plate-formes collaboratives : Lotus-Notes, Livelink, Hyperware, Microsoft
GED/Portail : Documentum, Vignette, ATG, Websphère, Microsoft, Plumtree
E-learning : Sum Total System, Saba, Hyperoffice, Syfadis, Innovae, Mindonsite
Autres : SAP, HR-Access, PeopleSoft

Produits :
L'offre « Knowledge Management » (KM) d'Unilog est gérée par Unilog Management. Elle consiste en l'étude et le déploiement de solutions spécifiques et intégrées pour ses clients dans les domaines de l'intelligence économique et de la gestion de la connaissance.
L'offre englobe :

. une composante conseil « métier » pour la définition fonctionnelle et l'accompagnement du déploiement du dispositif,
. une composante conseil « technologies » pour la conception de l'architecture du système et le choix des briques technologiques à intégrer (cf. partenariats),
. une composante développement.

VECSYS
www.vecsys.fr

- veille Internet
- **recherche et indexation**
- text mining
- data mining
- traduction
- traitement de l'image
- **traitement de la parole**
- représentation graphique
- knowledge management

Adresse :
ZA de Courtabœuf – Les Ulis – 3, rue de la Terre de Feu – 91952 COURTABŒUFCedex

Date de création : 1979

Partenariats :
Recherche : LIMSI (CNRS)
Autres partenariats : ANVAR – COMPAQ – CAP GEMINI ERNST & YOUNG – CSC – DIALOGIC – NATURAL Microsystem – RNRT – SEMA – STERIA – TENOR

Activité :
Le savoir-faire société provient du LIMSI -CNRS (laboratoire d'informatique pour la mécanique et les sciences de l'ingénieur).
2 secteurs d'activité :

- traitement automatique de la parole,
- produits et systèmes de communication associant les techniques de la micro-informatique avec les moyens de communications (radio, téléphone, réseaux sans fil, …) pour fournir une solution clef en main matériel et logiciel.

Produits :
OPENVOX : Systèmes de dialogue sur serveur téléphonique ou borne multimédia, intégrant des modules de reconnaissance de la parole continue, de synthèse de parole par concaténation d'unités enregistrées, et de dialogue et de compréhension du langage parlé. Le dialogue peut être en langage plus ou moins contraint selon que l'utilisateur peut être formé (applications professionnelles) ou non (grand public) et selon que l'on privilégie la fiabilité ou la flexibilité du dialogue.
2 applications :

- OPENVOX LN : serveurs vocaux en langage naturel,
- OPENVOX DG : serveurs vocaux en dialogues guidés.

MEDIASPEECH : Systèmes de transcription automatique de la parole spontanée, indépendamment du locuteur, pour diverses sources (radio, TV, web, …)

DATAVOX : Système de reconnaissance et synthèse vocale sur matériel autonome permettant les applications telles que commandes vocales dans l'automobile, simulateurs (avions, chars, contrôle aérien ,…), ou contrôle vocal domotique.

Partenariat avec le LIMSI – Laboratoire d'informatique pour la mécanique et les sciences de l'ingénieur (CNRS) - acteur majeur au niveau international.
VECSYS procède à un stockage de plusieurs mots différents tant par la langue que par la compréhension. Ces mots sont enregistrés entre autres à partir des chaînes de télévision par satellite et archivés. Ils constituent ainsi une bibliothèque permettant la transcription de la voix.

WEBFORMANCE www .webformancewatch.com	• **veille Internet** • **recherche et indexation** • **text mining** • data mining • traduction • traitement de l'image • traitement de la parole • représentation graphique • knowledge management	

Adresse :

9 Rue Maurice Grandcoing 94200 Ivry-sur-Seine

Date de création : 1996

Partenariats :
Sysqua, Yahoo, Ipea, Logilab

Concurrents **:**

Produits :
WebformanceWatch : plate-forme de veille permettant la collecte, le tri, le stockage, l'analyse et la diffusion des résultats.

WYSIGOT

Nom légal : MORELLE VINCENT (source : ABIL)

www

- **veille Internet**
- recherche et indexation
- **text mining**
- data mining
- traduction
- traitement de l'image
- représentation graphique
- knowledge management

Adresse :
La Tuilerie,33580 St Sulpice de Guilleragues

Date de création : 1998

Partenariats :

Concurrents **:**

Produits :
Wysigot : Capture et surveillance de sites

Le logiciel de Wysigot automatise la vérification des changements intervenant sur les sites web. Malheureusement, ses alarmes manquent parfois de lisibilité. L'avènement des forfaits de connexion a considérablement émoussé l'attrait pour les aspirateurs de sites web. Pourtant fondés sur des fonctions comparables, les logiciels de veille, dont Wysigot Plus fait partie, tirent leur épingle du jeu grâce à leur capacité de vérification des changements intervenus sur les sites.

Wysigot dispose d'assistants qui guident l'utilisateur travaillant avec Windows. Un bon point, les masques permettent de paramétrer le type de connexion ADSL ou LAN, et de préciser l'adresse du serveur proxy exploité par l'entreprise. En revanche, contrairement à son prédécesseur baptisé eCatch, Wysigot ne dispose pas de gestionnaire de bande passante pour éviter qu'un salarié mal intentionné ne s'accapare les lignes spécialisées de son entreprise. L'interface est sobre et reprend globalement l'agencement de l'explorateur Windows. Le plan de travail de gauche présente une structure arborescente des sites à surveiller. Cette structure est composée des répertoires des sites et de leurs pages. Le plan de droite offre diverses vues sur les documents web, telles que la liste des différentes profondeurs téléchargées, le statut des pages (type, modifié, à vérifier, taille, adresse, état) ou les pièces jointes constitutives du site (vidéo, programme, images, feuilles de style). Ces vues, qui permettent de contrôler les paramètres de surveillance d'une page, demeurent visualisables au moyen d'un navigateur web intégré.

Pour l'utilisateur, le principal attrait de Wysigot repose sur les paramètres de surveillance des sites et le réglage des alertes de modification. Pour chaque site ou document, la vérification des changements peut être déclenchée quelques jours après le premier téléchargement, au moment de la consultation hors ligne du document, à la demande ou en mode automatique. Suivant cette procédure, un algorithme lance le téléchargement des pages et des fichiers à des horaires différents, ceci afin d'économiser la bande passante. Dans cette même optique, il est possible d'inclure ou d'exclure de cette surveillance les textes, les images, les vidéos, les fichiers joints et même... les fichiers HTML.

Souple, Wysigot dispose d'options afin de régler la profondeur de l'aspiration, depuis la racine d'un site jusqu'à des niveaux d'arborescence plus profonds. Attention, cependant, le paramétrage de plus de quatre niveaux d'aspiration peut déboucher sur une diminution rapide de la bande disponible. Lorsqu'un changement intervient, le logiciel déclenche une alarme présentée sous forme d'une liste déroulante prévue à cet effet. Si cette modification est textuelle, la zone modifiée apparaît surlignée. Il est, par ailleurs, possible de paramétrer le logiciel afin que celui-ci imprime automatiquement toute page modifiée. Dans le but de prévenir une occupation excessive du disque dur par les fichiers téléchargés, Wysigot peut procéder au compactage des données et à leur stockage au sein du SGBD embarqué MSDE. Les utilisateurs qui désirent limiter l'accumulation de versions différentes d'un même site peuvent aussi en réduire le nombre. Tous les sites sont exportables selon leur format et leur structure d'origine. Autre bénéfice, Wysigot permet de s'authentifier automatiquement par un nom et un mot de passe.

Wisigot Plus est destiné aux spécialistes de la veille économique et financière, qui souhaitent identifier les pages web ayant fait l'objet d'un changement. En revanche, il est impossible de collecter automatiquement les mots à l'origine de ces modifications, dans le but de constituer une base de surveillance, par exemple. Wisigot ne convient pas à la surveillance de sites fortement transactionnels. Les alarmes trop nombreuses deviennent alors difficiles à exploiter, ceci en dépit d'une option censée prévenir ce problème

Vendu au prix de 38 euros.
Une version allégée dite « light » est disponible gratuitement.

XYLEME http://www.xyleme.fr/	• **veille Internet** • **recherche et indexation** • **text mining** • data mining • traduction • traitement de l'image • représentation graphique • knowledge management	

Adresse :
6 rue Emile Verhaeren 92 SAINT CLOUD

Date de création : 2000

Partenariats : TEMIS, Inxight, Sinequa, LTU technologies, UNISYS, Bull, Cap Gémini, Easypress, Go Albert

Concurrents :
Verity (US), Autonomy (GB), Oracle, Tamino et les moteurs de recherche en général

Produits :
Xyleme zone server : Indexation, classification, stockage, alertes automatisées. Spécialisé dans le traitement du texte semi-structuré. Outils sémantiques.
Facturation: exemple 100 utilisateurs, package de base, 1,5Go (uniquement de XML) = 70,000€

2ème partie : Recensement des laboratoires français

CNRS - INIST
Institut de l'information scientifique et technique
www.inist.fr

- veille Internet
- recherche et indexation
- **text mining**
- data mining
- traduction
- traitement de l'image
- représentation graphique
- knowledge management

Adresse :
2 allée du Parc de Brabois – 54514 Vandoeuvre-Lès-Nancy Cedex

Partenariats :
INSERM

Axes de recherche et produits :
Miriad : Plate-forme bibliométrique - Traitement linguistique
Plate-forme ILC : Traitement du langage naturel - Traitement linguistique
SDOC : Traitement infométrique (méthode des mots associés)

CNRS - LIMSI - TLP

(Laboratoire d'informatique pour la mécanique et les sciences de l'ingénieur)
Groupe TLP (Traitement du langage parlé)
www.limsi.fr/tlp/

- veille internet
- recherche et indexation
- text mining
- data mining
- traduction
- traitement de l'image
- **Traitement de la parole**
- représentation graphique
- knowledge management

Adresse:
BP 133, 91403 Orsay cedex

Date de création :
1972 pour le LIMSI
1995 pour le groupe TLP

Environ 120 permanents (chercheurs et enseignants) et 60 doctorants

Partenariats :
Vecsys, DGA/CTA, BBN, Cambridge University (CUED), RWTH, IRST, etc.

Concurrents :
Cambridge University (CUED), BBN, IBM, etc.

Axes de recherche et produits :
Les recherches du groupe visent à augmenter la compréhension des processus de la communication parlée et à développer des modèles appropriés au traitement automatique de la parole. Les problèmes scientifiques concernent aussi bien les modélisations acoustiques, lexicales et syntaxiques, que le lien entre parole et sens, ainsi que la modélisation des processus de communication. Ces problèmes, par essence pluridisciplinaires, nécessitent des compétences en traitement du signal, en acoustique, en phonétique, en linguistique et en informatique. Nos recherches nous amènent à développer des systèmes multilingues de traitement du langage parlé assurant des fonctions variées telles que la reconnaissance de la parole, l'identification de la langue et du locuteur, le dialogue oral homme-machine et l'indexation de documents audio et audiovisuels.

La reconnaissance de la parole consiste à convertir le signal audio en texte. Suivant l'usage visé, cette transcription peut être plus ou moins complète, avec le marquage des ponctuations, des hésitations et de certains évènements non linguistiques. La langue dans laquelle s'exprime le locuteur peut être identifiée en amont du système de reconnaissance lorsque celle-ci n'est pas connue a priori. L'identification du locuteur consiste à déterminer qui parle et quand, cette identification pouvant être absolue ou relative au document traité. La modélisation du dialogue oral dans les interfaces homme-machine va bien au delà de la transcription de la parole en texte, puisqu'il faut mettre en oeuvre des processus de compréhension et des stratégies de dialogue. Enfin, l'indexation automatique de documents audio pour l'accès à l'information par le contenu, nous amène à combiner les techniques de traitement de la parole et les techniques de traitement du langage naturel.

Le LIMSI développe des systèmes de traitement automatique pour ces différentes tâches, et en particulier pour la reconnaissance et l'indexation automatique d'émissions radio-télévisées (en anglais, français, allemand, espagnol, italien, portugais, arabe, mandarin) et de parole conversationnelle (en anglais, français, espagnol, arabe).

Intérêt :
Le LIMSI (groupe TLP) participe depuis 1992 aux évaluations internationales en reconnaissance de la parole (principalement conduites par le NIST). Il se classe systématiquement dans le peloton de tête et régulièrement premier. Les seuls autres laboratoires du même niveau sont l'université de Cambridge en Angleterre et quelques laboratoires américains (principalement BBN, SRI, IBM). De plus, ces performances élevées se retrouvent directement dans les produits valorisés par Vecsys.

Le groupe LIR participe également à des évaluations où il obtient de bons résultats, mais ne valorise pas pour l'instant ses travaux.

Valorisation de la technologie au travers de la société Vecsys.

CNRS - LIMSI - LIR Département Communication Homme-Machine Groupe Langue, Informations et Représentation	• veille internet • **recherche et indexation** • text mining • data mining • traduction • traitement de l'image • représentation graphique • knowledge management	

Adresse:
BP 133, 91403 Orsay cedex

Date de création du groupe : janvier 2001

Axes de recherche et produits :
Les activités de recherche sont essentiellement consacrées au traitement des données écrites, à leur analyse, leur compréhension et leur génération. Elles s'articulent autour de trois thèmes :

- Connaissances et raisonnement
- Documents : indexation, structuration, classification
- Processus d'analyse, génération et dialogue

Une action fédératrice sur le sujet Question/Réponse permet l'intégration, la mise en valeur et l'évaluation des travaux développés dans le groupe.
DOCUMENTS : INDEXATION, STRUCTURATION ET CLASSIFICATION :
La recherche d'information en est l'objectif central. Elle s'appuie sur des documents annotés, au plan morpho-syntaxique mais aussi sémantique.

Structuration et classification des documents : Le découpage en unités thématiques a été utilisé pour créer des classes sémantiques de noms pleins afin de permettre un enrichissement des requêtes en recherche documentaire.
La catégorisation de textes selon des types se révèle nécessaire si l'on veut utiliser efficacement des traitements tels que des étiqueteurs, des analyseurs et des systèmes de recherche d'information. En effet ceux-ci dépendent en particulier du « style » des textes, c'est-à-dire de la manière dont ils sont écrits. Des dispositifs sont développés pour obtenir des ensembles homogènes de données textuelles par le « style » et ainsi améliorer les traitements automatiques, en les spécialisant par apprentissage sur ces ensembles. Le profilage de textes utilise un traitement statistique multidimensionnel d'indices linguistiques (emploi du vocabulaire, de catégories morpho-syntaxiques, syntaxiques, sémantiques, structurelles et de patrons morpho-syntaxiques...) dans les parties d'un corpus multiplement annoté, pour regrouper ensuite ces parties en sous-ensembles homogènes sur ces points.
Indexation : Le travail en indexation textuelle au LIMSI s'est spécialisé dans l'extraction d'index multi-mots et dans la reconnaissance et le regroupement des variantes de ces index. L'analyseur FASTER permet de retrouver en corpus les principales variantes linguistiques d'index multi-mots en combinant analyse superficielle et exploitation de liens linguistiques (liens sémantiques et morphologiques). Des versions en sont développées pour le japonais, l'allemand, l'espagnol et le catalan. L'extraction et l'acquisition d'index ont été étendues aux noms propres et aux textes semi-structurés (HTML), pour acquérir et reconnaître les noms propres tels que des noms de lieux, de personnes ou d'institutions.

PROCESSUS D'ANALYSE, GENERATION ET DIALOGUE
Le dialogue homme-machine constitue le centre des activités de ce thème. Lui sont associés des processus d'analyse (pour comprendre les entrées) et de génération (pour engendrer les sorties). Ces activités d'analyse sont également développées pour la compréhension de textes.

Analyse et génération Plusieurs études sont en cours pour obtenir des analyses syntaxiques et sémantiques robustes. Un premier travail (contrat CIFRE avec Xerox) concerne le développement d'un analyseur syntaxique robuste capable de traiter avec une haute précision (au moins 96%) des corpus variés contenant des phénomènes très hétérogènes (corpus journalistiques, transcriptions de l'oral, rapports scientifiques divers, manuels techniques). Le second a pour point de départ les résultats d'un analyseur partiel et robuste (plusieurs ont été testés) et propose de compléter ces résultats partiels pour fournir une représentation sémantique sous la forme d'un ou plusieurs graphes conceptuels.

Sont également testés les résultats obtenus en utilisant des analyseurs syntaxiques (Sylex pour le français, Link Parser pour l'anglais) dans une tâche d'acquisition sémantique. Nous souhaitons regrouper ces expertises d'utilisation et de développement d'analyseurs robustes pour la réalisation de processus tels que la compréhension de textes et le dialogue homme-machine.

L'activité génération est centrée sur l'élaboration d'un outil d'aide à la rédaction.
Dialogue et annotation : Le groupe a développé un modèle de dialogue qui repose sur la théorie du discours et des Plans Partagés de Grosz et Sidner et des extensions apportées par Lochbaum. Le but est de permettre la modélisation de situations où la collaboration entre agents pour la réalisation d'une tâche commune est essentielle. Il faut pour cela modéliser l'ensemble des croyances et intentions que les agents d'un dialogue doivent avoir pour que leur collaboration puisse réaliser leur but commun.

ENST Bretagne – TAMCIC Laboratoire TAMCIC (Traitement algorithmique et matériel de la communication, de l'information et de la connaissance) – FRE 2658 Groupe I2RC (Intelligence des informations et réseaux de connaissance) http://www.enst-bretagne.fr/recherche/labo/labs/presentation.fr.php?idL=4	• veille internet • recherche et indexation • **text mining** • data mining • traduction • traitement de l'image • représentation graphique • knowledge management	

Adresse :
Technopôle Bres)I

Effectifs : environ 45 chercheurs pour le groupe.

Axes de recherche et produits :
Les activités de recherche de l'équipe concernent en particulier l'extraction terminologique, la classification automatique de documents, la constitution de lexiques à partir de documents électroniques. Sa spécificité est de pouvoir réaliser ces outils et ressources en contexte multilingue, notamment sur des langues « rares ».

Depuis 1986, le CRIM est à l'initiative de nombreux projets d'ingénierie linguistique, la plupart du temps en partenariat avec d'autres centres de recherche (Laboratoire d'Informatique de Paris 6, Dublin City University, etc.), des organismes institutionnels (Union Européenne, Organisation Mondiale de la Santé, Ministère de l'Industrie, etc.), des entreprises (EDF, Thales, Presses Universitaires de France, etc.).

Les projets du CRIM :
2002-2003
Princip : Détection automatique des contenus racistes et révisionnistes sur l'Internet
Lexiques : Création de lexiques trilingues (français anglais arabe) à partir de données électroniques en médias, relations internationales, langue des affaires
Evalda : Expertise et évaluation des méthodologies d'alignement de corpus parallèles
2001-2002
Safir : Création d'un méta-moteur de recherche sémantique dans le domaine des énergies renouvelables.

ENST - LTCI
Laboratoire LTCI (Traitement et Communication de l'Information)
UMR CNRS/GET Telecom Paris 5141

- veille internet
- **recherche et indexation**
- text mining
- data mining
- traduction
- traitement de l'image
- représentation graphique
- knowledge management

Adresse :
ENST / LTCI - 46, rue Barrault - 75634 PARIS Cedex 13

Date de création : 1982

Axes de recherche et produits :
Le LTCI poursuit 4 opérations de recherche (OR) :

OR *Communications, Électronique*
Ce domaine scientifique est l'intégration des systèmes de communications. Il regroupe l'ensemble des technologies vectrices de l'information multimédia dans les réseaux. Il comporte, à ce titre, les techniques, les concepts mathématiques, algorithmiques et physiques, les réalisations matérielles et les différents traitements mis en oeuvre pour le transport de l'information. Il a bénéficié pour cette période du développement des études sur l'Ultra Large Bande (UWB) ouvrant les problématiques MultInPut MultiOutPut (MIMO), antennes à large bande et impulsions ultra courtes, le développement des réseaux d'accès, le codage d'image, les besoins en sécurité avec la sécurité des circuits et la cryptographie optique, la compatibilité électromagnétique.

OR *Informatique et Réseaux*
Les axes de recherche des groupes du département Informatique et Réseaux sont :
- l'approfondissement et le renforcement de leur contribution en architecture de systèmes informatiques interconnectés dans un souci d'interopérabilité et de mobilité ;
- la définition et la réalisation d'outils permettant à la fois de concevoir, de valider, d'évaluer et de modéliser les architectures informatiques et les réseaux proposés tant d'un point de vue qualitatif (vérification) que quantitatif (évaluation) ;
- l'accentuation des travaux portant sur la représentation informatique (avec les traitements associés) d'un certain nombre de modes naturels de communication, et la modélisation informatique de processus cognitifs ;
- la définition et le déploiement de nouvelles architectures performantes de réseaux hétérogènes, fixes et mobiles avec des services de voix et de données pour intégrer les nouvelles complexités (taille, hétérogénéité, mobilité, sécurité, interopérabilité, services configurables, contenus riches) ;
- la compréhension des protocoles dans les couches basses pour la gestion optimale de la ressource radio en adéquation avec les protocoles associés et l'utilisation forte des protocoles logiciels pour vaincre la essource rebelle et rare de la radio par une intelligence protocolaire appropriée ;
- l'extension de méthodes et de boîtes à outils permettant le développement de très grands programmes utilisés dans des contextes où sont manipulées des données complexes (travail coopératif, multimédia ou facturation des services par exemple) ;
- la poursuite de travaux indispensables, issus des mathématiques appliquées et de l'informatique théorique, à des fins de modélisation et d'explicitation d'artefacts complexes (cryptographie, processus stochastiques, combinatoire, automates, ...).

OR *Traitement du Signal et des Images*

L'Opération de Recherche Traitement du Signal et des Images (OR TSI) est en charge au LTCI des recherches visant à améliorer l'extraction, le transport et la manipulation de l'information. Elle ancre ses travaux d'une part sur les mathématiques appliquées, d'autre part sur la physique et puise abondamment dans les sciences de l'homme et de la société pour étayer ses méthodes et orienter ses résultats. L'OR TSI est largement ouverte sur des domaines multiples liés aux applications de ses recherches : la santé, l'espace, le champ culturel, l'industrie du langage, la défense. Elle poursuit des collaborations étroites avec des organismes de recherche publique (Institut Géographique National, Centre National d'Etudes Spatiales, Assistance Publique des Hôpitaux de Paris, Centre Commun de Recherche des Musées de France, Centres Techniques de l'Armement, Institut National de l'Audiovisuel, Commissariat à l'Energie Atomique, ...) et privés (France Télécom R&D, Thalès, EADS, Alcatel, Motorola, ...).

OR *Économie, Gestion, Sciences Humaines et Sociales*

IMAG – CLIPS- GEOD (Institut d'Informatique et Mathématiques Appliquées de Grenoble) Laboratoire CLIPS (Communication Langagière et Interaction Personne-Système) – UMR CNRS/UJF/INPG 5524 Équipe GEOD (Groupe d'Étude sur l'Oral et le Dialogue)	• veille internet • **recherche et indexation** • text mining • data mining • traduction • traitement de l'image • représentation graphique • knowledge management	

Adresse :
GEOD - 220, rue de la chimie - Bât C - 1er Etage - B.P. 53 - 38041 Grenoble Cedex 9
http://www-clips.imag.fr/geod/

Axes de recherche et produits :

RAP : Reconnaissance Automatique de la Parole
L'équipe **GEOD** travaille sur la reconnaissance de parole continue grand vocabulaire du français, et sa mise en application dans des interfaces homme-homme ou homme-machine (systèmes de dialogue ou de traduction automatique de parole par exemple).
L'approche spécifique de l'équipe en modélisation du langage consiste à utiliser Internet pour apprendre les modèles de langage. Les autres activités sont liées à la reconnaissance automatique distribuée et les serveurs vocaux. En effet, il est de plus en plus fréquent de trouver des systèmes de reconnaissance de parole dans les téléphones mobiles ou dans d'autres terminaux de communication. Si la reconnaissance de mots clés peut être faite en local sur le terminal, la reconnaissance de parole continue grand vocabulaire nécessite un traitement complexe qu'il n'est pas encore possible de réaliser sur des terminaux clients standards. Dans ce cas, le signal de parole doit être acheminé du terminal client, vers le serveur distant qui effectuera les traitements nécessaires à la reconnaissance. Le signal de parole est le plus souvent codé pour diminuer les temps de transmission. Ce codage peut cependant introduire des distorsions du signal de parole qui ont pour conséquence de dégrader les performances du serveur de reconnaissance.

RLA : Ressources Linguistiques et Apprentissage
Les ressources linguistiques constituent le matériau indispensable pour le développement et l'optimisation des systèmes de reconnaissance automatique de la parole. L'élaboration de ces ressources relève de recherches transversales qui concernent aussi bien le développement d'outils ou de méthodologies pour l'acquisition et la gestion des corpus de parole que le développement ou l'adaptation de techniques d'apprentissage et de recherche d'information.
L'objectif visé est l'étiquetage automatique (à divers niveaux : phonétique, syntaxique, sémantique, thématique.) de grandes bases de données de parole, mais aussi, d'une façon plus générale, l'indexation des bases de données sonores.

RRA: Robustesse dans la reconnaissance et l'acquisition du signal de parole
Le problème majeur de la reconnaissance de la parole est son utilisabilité dans des conditions réelles : les systèmes ne sont pas à l'heure actuelle suffisamment robustes aux inattendus de la langue orale ni aux conditions acoustiques sévères. Cela impose une amélioration des systèmes pour prendre en compte la parole spontanée accompagnée de tous les phénomènes de variabilité dus aux différents locuteurs et aux environnements instables.
La reconnaissance robuste de la parole, pour pouvoir être efficace dans des domaines d'utilisation les plus vastes possibles, doit s'appuyer sur toutes les ressources disponibles dans l'environnement, l'application (ou la situation) et le langage.

Par ailleurs pour des applications réalistes d'interaction homme-machine, il est indispensable de fonctionner en "temps réel".
C'est pourquoi, un ensemble de modules acoustiques de prétraitement du signal, de rehaussement de la parole (méthode de séparation de sources), de filtrage des signaux sonores en provenance de l'environnement et nuisibles à la parole, permet d'améliorer le rapport signal/bruit pour augmenter les performances du système de reconnaissance.

DIM: Dialogue oral homme-machine & Interfaces Multimodales
Le thème du dialogue homme-machine englobe l'interaction orale et l'interaction multimodale (parole et gestes). La modélisation du dialogue homme-machine pose des problèmes théoriques car le dialogue humain ne peut être considéré comme une activité entièrement planifiée : à chaque instant les interlocuteurs peuvent opérer des incidences ou des ruptures ils utilisent des stratégies qu'ils adaptent au cours de l'interaction en fonction des buts à atteindre et des opportunités offertes par la situation.
Le dialogue homme-machine n'a d'utilité que dans un cadre opératoire, dit finalisé, c'est-à-dire pour effectuer des tâches coordonnées (résoudre des problèmes, renseigner, aider à la conception, assister l'enseignement, etc.), ce qui place le dialogue dans une relation opérateur/tâche où la machine a un rôle collaboratif.
A travers et par le dialogue, la machine doit également apprendre de nouvelles actions ou optimiser son comportement face à de nouvelles situations. Pour cela, elle doit pouvoir inférer et gérer les buts de l'utilisateur, comprendre ses actes de langage, être capable de les interpréter en fonction de la situation pour finalement générer et effectuer l'action ou le plan d'action le plus adéquat.
La conception de modèles de dialogue et la réalisation de systèmes temps réel est l'objectif de l'équipe.

IMAG – CLIPS - MRIM
(Institut d'Informatique et Mathématiques Appliquées de Grenoble)
Laboratoire CLIPS (Communication Langagière et Interaction Personne-Système) – UMR CNRS/UJF/INPG 5524
Équipe MRIM (Modélisation et Recherche Multimédia)
http://www-mrim.imag.fr/

- veille internet
- **recherche et indexation**
- text mining
- data mining
- traduction
- traitement de l'image
- représentation graphique
- knowledge management

Adresse :
MRIM - Bâtiment B - CLIPS - IMAG - BP 53, 38041 Grenoble Cedex 9

Axes de recherche et produits :

L'équipe a pour objectif de proposer des systèmes de recherche d'information opérationnels. Elle a développé des outils d'indexation de texte (le projet IOTA) et un environnement d'indexation assistée d'images et est en train de constituer les éléments de base pour l'indexation de la vidéo.

- Systèmes de recherche de documents textuels.
- Systèmes de recherche d'images.
- Systèmes de recherche de vidéo.

Systèmes de recherche de documents textuels
Axe de recherche : RITM (Recherche d'Information Textuelle et Multilingue)
Le prototype IOTA fournit une indexation automatique du texte intégral, une construction automatique du thesaurus du domaine. Le vocabulaire étant complètement ouvert, le langage d'indexation est basé sur l'extraction de syntagmes nominaux. Un analyseur syntaxique de surface dédié à la recherche d'information a été défini et implanté. Le premier prototype IOTA a été développé dans les années 90, nous l'avons redéfini et consolidé afin qu'il puisse prendre en compte des corpus en vraie grandeur de plusieurs Giga-octets. Sa robustesse est à l'heure actuelle suffisante pour nous permettre de participer à des actions de tests internationales comme TREC. Un autre aspect sur lequel nous travaillons est l'extraction automatique de la connaissance (ou thesaurus), d'un point de vue pratique et théorique. Nous participons à l'action AUPELF UREF Arc A1 et A3 qui est dans ce sens très significative, car elle permet de valider d'un point de vue qualitatif l'extraction de la terminologie d'un domaine et son efficacité en terme de rappel est précision. L'intégration de l'indexation à base d'arbres syntagmatiques est en cours.

Systèmes de recherche d'images
Axe de recherche : IF (Images Fixes)
Le prototype RIME permet l'indexation, la navigation et l'interrogation d'un corpus structuré et multimédia : le corpus traité est un corpus médical contenant des données administratives de patients, des comptes rendus médicaux et des images médicales. L'objectif de ce prototype est de permettre d'indexer finement les données (textes, images médicales, dossiers médicaux) afin de les retrouver par leur contenu. Le langage d'indexation de toutes ces données (textes et images) est un langage complexe permettant une recherche d'informations orientée précision : la précision étant la qualité essentielle dans un SRI destiné à des utilisateurs spécialistes (les médecins). Le langage d'indexation est basé sur le formalisme des graphes conceptuels (le modèle des Graphes Conceptuels de Sowa).
Une plate forme complète de définition et de manipulation des graphes conceptuels a été implantée sur le SGBD à objets O2. Cette plate forme évolue actuellement selon deux axes : l'intégration des résultats théoriques obtenus, d'une part, et l'optimisation du point de vue efficacité des algorithmes de correspondance des graphes conceptuels, d'autre part.

La construction d'index basée sur ce langage d'indexation nécessite une analyse des documents assistée par l'indexeur. D'un point de vue pratique, les données et leurs index sont stockés sur le SGBD O2, l'interrogation et la navigation sont accessibles sur le WWW, en utilisant la passerelle O2Web. Le modèle de données multimédia développé par l'équipe est appliqué à RIME. Ce prototype a servi de plate forme de tests pour le projet FERMI, la base d'images constituée de 650 images de Paris a été indexée en utilisant le langage d'indexation d'images et l'on peut rechercher des images selon plusieurs facettes.

Systèmes de recherche de vidéo
Axe de recherche : V (Vidéo)

Pour l'instant, nous n'avons pas un SRI complet. Nous mettons en un système d'indexation assistée et de recherche par le contenu de documents vidéos. Ce système est basé sur une indexation automatique partielle dont les résultats peuvent être corrigés ou complétés par l'utilisateur en fonction du compromis coût qualité recherché. Plusieurs briques de base de ce système sont développées ou en cours de développement (segmentation en plans, recherche des mouvements de caméras, détection et suivi d'objets mobiles, caractérisation par vecteurs de caractéristiques, détection de personnages). parallèlement à ces traitements effectués sur la bande image, une indexation à partir de la transcription de la bande son ou des sous-titres lorsqu'ils sont disponibles sera intégrée et couplée à un système de détection de thèmes par analyse de la terminologie (en relation avec le projet THEOREME)

IMAG – CLIPS - GETA
(Institut d'Informatique et Mathématiques Appliquées de Grenoble)
Laboratoire CLIPS (Communication Langagière et Interaction Personne-Système) – UMR CNRS/UJF/INPG 5524
Equipe GETA (Groupe d'Etude pour la Traduction Automatique)
http://www-clips.imag.fr/geta/

- **veille internet**
- recherche et indexation
- **text mining**
- data mining
- **traduction**
- traitement de l'image
- représentation graphique
- knowledge management

Adresse :
385, rue de la Bibliothèque - B.P. 53 - 38041 Grenoble Cedex 9

Date de création : 1971 (issu du CETA 1961-1971)

Axes de recherche et produits :

Traduction assistée par ordinateur pour le rédacteur et pour le traducteur
Traitement multilingue de l'information

- Ariane-Y, unifier et étendre l'environnement de développement de systèmes de TAO du GETA;
- C-STAR projet international pour l'évaluation de la traduction de parole; http://www.c-star.org
- NESPOLE!, projet européen de traduction de la parole en quatre langues sur Internet; http://nespole.itc.it
- PAPILLON, construction coopérative d'une base lexicale multilingue comprenant l'anglais, le français, le japonais, le malais, le lao, le thai et le vietnamien; http://www.papillon-dictionary.org
- UNL, communication et recherche d'information multilingue sur le réseau; http://www.unl.ias.unu.edu

INALCO Laboratoire CRIM (Centre de recherche en ingénierie multilingue) http://www.inalco.fr/ina_gabarit_rubrique.php3?id_rubrique=1531	• veille internet • **recherche et indexation** • **text mining** • data mining • traduction • traitement de l'image • représentation graphique • knowledge management	

Adresse :
2, rue de Lille - 75007 Paris

Axes de recherche et produits :
Les activités de recherche de l'équipe concernent en particulier l'extraction terminologique, la classification automatique de documents, la constitution de lexiques à partir de documents électroniques. Sa spécificité est de pouvoir réaliser ces outils et ressources en contexte multilingue, notamment sur des langues « rares ».

Depuis 1986, le CRIM est à l'initiative de nombreux projets d'ingénierie linguistique, la plupart du temps en partenariat avec d'autres centres de recherche (Laboratoire d'Informatique de Paris 6, Dublin City University, etc.), des organismes institutionnels (Union Européenne, Organisation Mondiale de la Santé, Ministère de l'Industrie, etc.), des entreprises (EDF, Thales, Presses Universitaires de France, etc.).

Les projets du CRIM :
2002-2003
Princip : Détection automatique des contenus racistes et révisionnistes sur l'Internet
Lexiques : Création de lexiques trilingues (français anglais arabe) à partir de données électroniques en médias, relations internationales, langue des affaires
Evalda : Expertise et évaluation des méthodologies d'alignement de corpus parallèles
2001-2002
Safir : Création d'un méta-moteur de recherche sémantique dans le domaine des énergies renouvelables

INRIA - IMEDIA

(Institut National de Recherche en Informatique et Automatique)
www.inria.fr
Projet IMEDIA
http://www.inria.fr/recherche/equipes/imedia.fr.html

- veille Internet
- **recherche et indexation**
- text mining
- data mining
- traduction
- **traitement de l'image**
- représentation graphique
- knowledge management

Adresse :
Domaine de Voluceau - Rocquencourt
BP 105, 78153 Le Chesnay Cedex

Date de création :
INRIA : 1967 - 900 pers (400 chercheurs, 500 ingénieurs et techniciens)
IMEDIA : 1999 - 10 pers

Axes de recherche et produits :

L'objectif premier de l'équipe **IMEDIA** est de développer des méthodes d'indexation par le contenu, de recherche interactive, et de navigation dans des bases d'images, dans un contexte multimédia. Pour ce faire, nous traiterons aussi bien des bases d'images " génériques " (ex: web) que des bases d'images " spécifiques " à un domaine d'application ciblé (visages, images médicales...). Ces deux catégories relèvent respectivement de la recherche d'images et de la reconnaissance d'objets. En réalité la recherche d'images est une problématique plus vaste qui englobe la reconnaissance d'objets et intègre les intéractions avec l'utilisateur. Plus généralement, l'équipe IMEDIA déploie ses efforts de recherche, de collaboration et de transfert pour répondre au problème complexe de l'accès intelligent aux données multimédias dans sa globalité. 'interface utilisateur **IKONA** permet la recherche d'images par contenu visuel à partir de requêtes pertinentes.

MAESTRO est un moteur de recherche d'images par similarités visuelles.
Il est utlisé dans la recherche de motifs ou d'objets dans des bases d'objets d'art ou dans la recherche de logo dans une image.

Thèmes
- Indexation d'images par le contenu. Choix d'un espace de représentation et calcul des attributs significatifs des images, construction d'index (compromis coût du stockage/coût de la requête).
- Recherche interactive dans des grandes bases d'images. Similarité perceptuelle, mise en correspondance, modélisation de l'incertitude, profil utilisateur, contrôle de pertinence, requêtes partielles (sur des zones d'intérêt).
- Navigation. Théorie de l'information, clustering, modélisation probabiliste, résumé visuel, recherche d'image mentale, interfaces.

Indexation multimedia. Généralisation aux données hétérogènes, Hybridation texte-image, application aux documents multimédias.

Atouts technologiques
Brevets : dépôts des prototypes de logiciels à l'Agence de Protection des Programmes (APP).

Secteur ciblé :
Applications de sécurité (police), audiovisuel, applications scientifiques, culture et éducation, télécommunications.

INSA Lyon LIRIS - Laboratoire d'InfoRmatique en Images et Systèmes d'information FRE 2672 http://liris.cnrs.fr/	• veille internet • **recherche et indexation** • text mining • data mining • traduction • traitement de l'image • représentation graphique • knowledge management	

Adresse :
INSA de Lyon
Blaise PASCAL
69 621 VILLEURBANNE CEDEX

Axes de recherche et produits :
- Données, documents et connaissances (D2C)
- Images et vidéos : segmentation et extraction d'information (IV)
- Modélisation et réalité augmentée
- Systèmes d'information communicant

IGM - UMLV Institut Gaspard Monge – Université de Marne la Vallée Laboratoire d'Informatique Équipe informatique linguistique http://www-igm.univ-mlv.fr/LabInfo/equipe/linguistique/	• veille internet • **recherche et indexation** • **text mining** • data mining • traduction • traitement de l'image • représentation graphique • knowledge management • **traitement de la parole**	

Adresse :
77454 Marne-la-Vallée Cedex 2

Axes de recherche et produits :

L'équipe d'informatique linguistique se situe dans la perspective du traitement automatique des textes en langues naturelles. Elle privilégie l'utilisation de données précises et explicites : dictionnaires, grammaires, par rapport à l'approximation à partir de données incertaines. Les applications visées sont nombreuses mais les plus significatives sont liées à la recherche documentaire. Les trois orientations principales sont :

- la production de dictionnaires électroniques d'autres langues que celles pour lesquelles on dispose déjà d'outils fiables ;
- les traitements intermédiaires entre l'analyse lexicale et l'analyse syntaxique en vue de l'accès aux informations dans les grandes bases de textes : reconnaissance et indexation de terminologie, levée d'ambiguïtés...
- à plus long terme, la poursuite de la constitution des lexiques-grammaires, qui consistent en une description systématique et formelle de la syntaxe de langues naturelles.

Le niveau de couverture lexicale visé est très large, car rien n'indique a priori dans quels domaines se situent les textes susceptibles d'être traités par les applications ; de plus, nous nous intéressons aux applications dans lesquelles on n'impose aucune restriction de vocabulaire aux textes traités, qui sont entièrement libres, c'est-à-dire qui obéissent aux seules contraintes de la langue elle-même. La prise en compte du lexique est donc systématique.

IRISA
Laboratoire commun INRIA, CNRS, Université de Rennes 1, INSA Rennes
http://www.irisa.fr

- veille internet
- **recherche et indexation**
- text mining
- data mining
- traduction
- **traitement de l'image**
- représentation graphique
- knowledge management

Adresse :
Irisa/Inria de Rennes, Campus de Beaulieu 35042 Rennes Cedex

Date de création : 1973
Effectifs : 500 personnes (IRISA)

Axes de recherche et produits :

L'IRISA mène plusieurs projets de recherche et développement technologique dont les retombées peuvent intéresser le domaine des outils de veille, en particulier :

- *Projet **CORDIAL** : dialogue oral homme-machine*

Les activités du projet se déclinent sur deux axes complémentaires qui visent une mise en oeuvre facilitée de systèmes oraux interactifs conviviaux. Le premier axe est constitué d'études fondamentales dont les résultats sont considérés comme des préalables à la construction automatique de constituants de systèmes. Le second axe concerne les aspects pratiques de mise au point, d'intégration et d'évaluation de systèmes oraux interactifs.
Les études fondamentales concernent la modélisation de différents aspects de la communication homme-machine : modèles de dialogue - structural, intentionnel -, traitement de la référence, traitement des erreurs de communication et dialogue de médiation. Elles abordent aussi l'apprentissage automatique de structures syntaxiques, prosodiques et "dialogiques" dont le but à terme est de simplifier la mise en place de nouveaux systèmes. Enfin, des études sur la prosodie sont développées dans une optique pédagogique et d'amélioration de la synthèse de la parole.
Dans le cadre d'études de développement, nous avons mis au point plusieurs prototypes de systèmes oraux interactifs (interrogation multimodale de base de données, progiciel d'enseignement). Le difficile problème de l'évaluation des systèmes de dialogue est également traité.

- *Équipe **IMADOC** : interprétation et reconnaissance d'images et de documents*

Les recherches menées au sein du projet IMADOC (IMAges et DOCuments) concernent l'écrit et le document sous toutes leurs formes (manuscrit, imprimé, image, graphique, multimédia, etc.) ainsi que les activités qui y sont liées, notamment la production de nouveaux documents, la transformation sous forme électronique élaborée de documents papier existants et leur traitement « intelligent » ainsi que l'Interaction Homme-Document (I.H.D.). De manière plus générale, les centres d'intérêt du projet IMADOC touchent à la communication écrite sous un triple aspect : synthèse de documents, analyse de documents, interaction homme-document.

Axes de recherche :

- Reconnaissance de l'écriture manuscrite en-ligne et hors-ligne ;
- Extraction de la structure de documents (séparation de connaissance et traitement, grammaire) ;
- Communication Homme-Machine orientée stylo (modélisation de gestes, interface générique) ;
- Modèles et systèmes de perception (cycles perceptifs, introduction du contexte) ;
- Comparaison et recherche approchée de textes (recherche d'éléments communs à deux textes).

- *Projet **TEMICS** : traitement, modélisation et communication d'images numériques*

Les objectifs du projet sont de développer les concepts et les outils d'analyse, de modélisation, de codage, et de tatouage d'images, et plus généralement des informations vidèo manipulées en communication multimédia. Nos travaux portent plus particulièrement sur les problèmes suivants:

- l'interaction avec le contenu et la navigation dans des scènes vidéo 3D;
- la représentation compacte et robuste aux bruits de transmission des images et des signaux vidéo;
- le marquage (ou tatouage) des images et des signaux vidéo à des fins de protection contre les copies illicites, et à des fins d'authentification.

- *Projet **VISTA** : Vision spatio-temporelle et apprentissage*

Le projet VISTA s'intéresse à plusieurs types d'imageries spatio-temporelles relavant essentiellement de l'imagerie optique (vidéo, infrarouge), mais éventuellement aussi de l'acoustique (sonar, échographie). Nos travaux de recherche visent à l'analyse de scènes dynamiques, ou, plus généralement de phénomènes dynamiques, à partir de séquences d'images. Nous abordons l'ensemble des problèmes liés au traitement de ces contenus dynamiques et plus particulièrement à l'analyse du mouvement : détection, mesure, segmentation, suivi, reconnaissance, interprétation avec apprentissage. Nous privilégions une approche statistique de ces problèmes : modèles markoviens, inférence bayésienne, estimation robuste, filtrage particulaire, apprentissage statistique. Nous nous intéressons à quatre grands domaines d'applications :

- traitement et indexation vidéo
- imagerie météorologique et visualisation expérimentale en mécanique des fluides
- imagerie biologique
- surveillance et navigation

IRISA - SIAMES
Laboratoire commun INRIA, CNRS, Université de Rennes 1, INSA Rennes
Projet SIAMES (Synthèse d'Images, Animation, Modélisation et Simulation)
http://www.irisa.fr/siames/

- veille internet
- recherche et indexation
- text mining
- data mining
- traduction
- **traitement de l'image**
- représentation graphique
- knowledge management

Adresse :
Irisa/Inria de Rennes, Campus de Beaulieu 35042 Rennes Cedex

Effectifs : 500 personnes (IRISA)

Axes de recherche et produits :

Le projet aborde l'ensemble des méthodes nécessaires à la production de séquences d'images de synthèse. Trois axes sont principalement concernés :

- l'informatique graphique : où l'essentiel des travaux consiste à élaborer et intégrer des modèles, à définir des algorithmes et à étudier les complexités des solutions proposées ;
- la simulation : notre objectif principal est de pouvoir confronter les résultats produits par nos algorithmes à des valeurs numériques mesurées sur site réel, ceci afin de valider expérimentalement les approches et concepts étudiés ;
- l'organisation système : pour développer les deux points précédents, nous devons être à même de traiter des cas grandeur nature et valider nos approches par des mises en oeuvre.

Siames est un projet commun avec le CNRS, l'Université de Rennes 1 et l'Insa de Rennes.

Axes de recherche

- Modélisation de scènes tridimensionnelles : nous nous intéressons aux méthodes permettant de spécifier la géométrie des scènes (3D) complexes et, plus particulièrement, au développement de méthodes déclaratives (haut niveau) pour la description de la forme des objets en termes de contraintes et de propriétés.
- Simulation d'éclairage : les algorithmes de synthèse d'image réalistes permettent d'obtenir des résultats de très haute qualité par l'introduction de modèles d'éclairement fondés sur la physique, afin d'évaluer les interactions entre la lumière et les objets.
- Simulation/animation de systèmes physiques : nous abordons la simulation de systèmes physiques sous l'angle des schémas de calcul nécessaires pour la production des équations régissant ces systèmes. Nous étudions aussi la résolution de ce système d'équations (approche symbolique/numérique). Cette approche nous permet d'aborder les problèmes de simulation ou d'animation par ordinateur. Un autre point fondamental que nous étudions est le problème du contrôle du mouvement ainsi que la conception de systèmes d'animation (intégration des différents types de modèles physiques, plateforme d'animation/simulation, etc.).
- Algorithmes parallèles (thème transversal) : la tendance actuelle est largement orientée vers l'utilisation de modèles de plus en plus complexes (forme, mouvement, rendu). Les conséquences directes en sont la forte augmentation des coûts de calcul dûs à la production d'images fixes ou animées. Outre les recherches visant à réduire la complexité des algorithmes séquentiels, l'étude des schémas de parallélisation de ces algorithmes revêt un caractère fondamental. Ces travaux sont menés en étroite collaboration avec T. Priol du projet Caps.

IRIT - SIG

(Institut de Recherche en Informatique de Toulouse).
Équipe SIG (Système d'informations généralisées)
www.irit.fr
http://www.irit.fr/recherches/IRI/SIG/

- **veille Internet**
- **recherche et indexation**
- text mining
- data mining
- traduction
- traitement de l'image
- **représentation graphique**
- knowledge management

Adresse :
118 route de Narbonne 31062 TOULOUSE Cedex4

Date de création : janvier 1990
Habilitation : non
Effectifs : 200 personnes.
Sous tutelle du CNRS,
de l'Institut National Polytechnique de Toulouse, de l'Université Paul Sabatier.

SIG : une vingtaine de chercheurs

Partenariats :
Laboratoire commun MIDI, INTUILAB, Cryptomthic, CS systèmes d'information, EBS Toulouse, Microelectronics, Sinequa, Airbus, Alcatel Espace, Cnes, Eads, France Telecom, IBM France, Saint Gobain, SUN Microsystems, Thalès,..

Axes de recherche:
Les recherches menées au sein de l'équipe SIG s'articulent actuellement autour de concepts de Base d'Informations selon deux approches complémentaires permettant d'appréhender :

- les gisements d'informations et bases de données multidimensionnelles, dont l'objectif est de proposer, développer et expérimenter des techniques et stratégies d'identification de sources d'information, de mémorisation, de filtrage/recherche d'information (explicite ou cachée) et de présentation, et de coordonner ces opérations à travers un plan stratégique global.
- les systèmes d'information et d'ingénierie documentaire pour proposer, développer et expérimenter des modèles, des langages, des méthodes et des techniques autour du concept de bases d'objets documentaires ou hyperbases. Cette approche privilégie l'élicitation et la manipulation de structures irrégulières issues des informations manipulées, via des langages de type SQL, OQL, XmlQL,...

Ces deux axes reposent sur trois thèmes :

- ***Filtrage, Recherche, Exploration d'informations***

Cet axe concerne la mise au point de SRI (Systèmes de Recherche d'Informations) sophistiqués orientés bases textuelles avec la prise en compte du concept de profil (profil utilisateur, profil d'usage), l'intégration de possibilités d'interrogation multilingue (combinant par exemple Français, Anglais, Allemand,...), la reformulation automatique de requêtes par réinjection d'informations prenant en compte des préférences utilisateurs ou extraites via des outils d'analyses multidimentionnelles.

Cet axe repose sur deux composantes :

- SIG/RFI (Recherche et Filtrage d'Information)
- SIG/EVI (Extraction et Visualisation d'Informations

- ***Entrepôts de données (SIG/ED)***

Un entrepôt de données stocke des données utiles aux décideurs pour effectuer diverses analyses économiques grâce aux techniques de type On-Line Analytical Processing (OLAP) ou fouille de données (Data Mining). L'étude des modèles de description des entrepôts de données et des métas donnés permet de caractériser les données issues de sources hétérogènes et de prendre en compte leur historisation à différents niveaux de granularité. On apporte ainsi une extension aux concepts présents dans les Bases de Données Temporelles, on y appréhende les aspects Méthodologie de Conception d'entrepôts et de manipulation de données multidimentionnelles.

- ***Documents, Données Semi-Structurées et usages (SIG/DDSS)***

Actuellement, le concept de méta données (sprécifiées à priori , extraites ou générées) permet de décrire par média les ensembles d'informations constituant une base de données multimédia (description par facettes, par annotations qui peuvent être hiérarchisées, stratifiées, multidimentionnelles). A partir de ces descriptions orientées média, nous proposons une démarche d'unification pour en proposer une vision générique et permettre un traitement uniforme des requêtes qui vont combiner ces médias

Produits :

Tetralogie : logiciel de veille scientifique. Cet outil est un des éléments essentiels de la station bibliométrique "ATLAS" élaborée conjointement par le CEDOCAR et le SGDN.

LORIA UMR INRIA/CNRS/INPL/U. Nancy I et II 7503 http://www.loria.fr/index.php	• veille internet • **recherche et indexation** • **text mining** • **data mining** • traduction • traitement de l'image • représentation graphique • **knowledge management** • **traitement de la parole**	

Adresse :
Campus Scientifique - BP 239
54506 VANDOEUVRE-lès-NANCY CEDEX

Axes de recherche et produits :

Le LORIA est un laboratoire qui regroupe 25 équipes de recherche dont 6 intéressent directement le secteur de la veille :

- LANGUE ET DIALOGUE : Informatique linguistique pour le dialogue homme-machine multimodal
- ORPAILLEUR : Représentations de connaissances, raisonnements, et extraction de connaissances à partir de bases de données
- PARAOLE : Analyse, Perception et Reconnaissance automatique de la parole
- QGAR : Navigation dans les documents graphiques par l'analyse et la reconnaissance
- READ **:** Reconnaissance de l'écriture et analyse de documents
- SITE **:** Modélisation et Développement de Systèmes d'Intelligence Économique

Université d'Angers - ISTIA ISTIA (Institut des Sciences et Techniques de l'Ingénieur d'Angers) Laboratoire CPNI (Conception de Produits Nouveaux et Innovants) http://www.istia.univ-angers.fr/Innovation/pres-cpni.php3	• veille internet • **recherche et indexation** • text mining • data mining • traduction • traitement de l'image • représentation graphique • knowledge management	

Adresse :
62, avenue Notre Dame du Lac 49 000 Angers

Axes de recherche et produits :

3 axes de recherche :

- Méthodologies de conception et technologies de la réalité virtuelle,
- Cognitique et réalité virtuelle,
- Créativité, simulation mécanique et réalité virtuelle.

La Réalité Virtuelle est une technologie de pointe au croisement des sciences de la mécanique, de l'automatisme et de l'informatique. Son but est de reproduire en images de synthèse des objets ou des environnements et de permettre aux utilisateurs d'interagir en temps réel avec ces objets et ces environnements de manière intuitive. La Réalité Virtuelle modifie les environnements de travail collaboratif des équipes projets. Elle permet, d'une part, une diminution du temps de cycle de conception d'un produit en intensifiant les échanges (« ingénierie concourante », « travail collaboratif », «TIC »). D'autre part, elle intègre l'utilisateur final dans les phases initiales du processus de conception des produits industriels. Cependant, elle nécessite des systèmes informatiques capables d'optimiser le traitement de volumes croissants de données. Cette organisation de l'entreprise autour d'un échange d'information numérique est nommée "la chaîne numérique de la conception".
L'objectif de nos recherches est d'évaluer le potentiel des technologies de la réalité virtuelle et de modéliser le processus de conception d'un produit (objets, acteurs, raisonnements, méthodes, organisations) ainsi que les systèmes d'informations (SI) associés

Université d'Avignon – LIA - TALNE
LIA (Laboratoire d'Informatique d'Avignon) UPRES 921
Équipe Traitement automatique du langage naturel écrit
http://www.lia.univ-avignon.fr/equipes/TALNE/index.html

- veille internet
- **recherche et indexation**
- **text mining**
- data mining
- traduction
- traitement de l'image
- représentation graphique
- knowledge management
- traitement de la parole

Adresse :
339, chemin des Meinajaries, Agroparc BP 1228 - 84911 AVIGNON Cedex 9

Effectifs : 10 chercheurs

Partenariats industriels privilégiés :
Sinequa
Bertin
Memodata

Axes de recherche et produits :

Recherche Documentaire
Le partenariat bertin-lia a conduit au développement d'un ensemble de composants logiciels optimisés pour la recherche d'information dans les grandes bases de données textuelles.
Parallèlement au partenariat avec bertin technologies, le lia a développé son propre système de recherche documentaire : le système siac. Un prototype de ce système a été construit en 1997, année où il a participé à la première campagne d'évaluation Amaryllis. En 1998, siac a été utilisé dans la campagne trec-7 pour des expériences en classification automatique en collaboration avec le système Indexal de recherche conçu conjointement par bertin et le lia. En 1999, siac a participé à la deuxième campagne Amaryllis. Des progrès sensibles ont été réalisés par ce système entre la 1ère et la 2ème campagne Amaryllis en termes de précision et de rappel. Depuis la première version du système, des modules de classification et de segmentation thématiques par arbres de décision non supervisés ont été rajoutés comme décrit plus loin.

Lexique sémantique
Le lia a joué un rôle décisif pour que la langue française soit incluse dans le projet européen EuroWordNet (ewn). L'objectif du projet européen ewn était de réaliser un lexique sémantique multilingue du même type que WordNet (**http://www.cogsci.princeton.edu:80/~wn/**). La première phase du projet ne concernait pas toutes les langues majeures parlées en Europe. Seules étaient pris en compte l'italien, l'espagnol, l'anglais et le hollandais. La deuxième phase du projet a débuté en avril 1998, a duré quinze mois et concernait le français, l'allemand, l'estonien et le tchèque. Pour la partie française, en dehors du lia qui en était responsable, ce contrat a réuni trois partenaires : Xerox Research Centre Europe (xrce), Bertin Technologies et MemoData.

Désambiguïsation Sémantique

Privilégiant, comme en traitement de la parole, une combinaison d'approches numériques et symboliques, le lia a choisi d'étendre l'application de ce type de stratégies mixtes à d'autres niveaux que ceux concernés par le seul étiquetage syntaxique. Le système probabiliste ecsta d'étiquetage (syntaxique et morphologique) a été généralisé pour pouvoir prendre en charge la désambiguïsation du niveau sémantique au moyen de modèles de Markov. La difficulté du problème nous a conduit à diversifier les approches et à puiser dans la palette de techniques habituellement mises en œuvre en reconnaissance des formes : arbres de décision, recherche des plus proches voisins, règle de décision bayésienne, etc.

Des tests ont été effectués sur WordNet et le SemCor (corpus sémantique étiqueté à l'aide des sens WordNet) en utilisant des chaînes de Markov Cachées. Ces expériences ont montré l'intérêt de conserver plusieurs sens possibles pour un même mot plutôt que de n'en choisir qu'un seul. En effet, en utilisant l'algorithme de Baum-Welch, le taux de bon étiquetage est de 95,3 % parmi les 3 premières étiquettes choisies par le système alors qu'avec un algorithme de Viterbi, le taux de bon étiquetage pour l'étiquette choisie est de 72,3 %.

Université d'Avignon – LIA - TALNO LIA (Laboratoire d'Informatique d'Avignon) UPRES 921 Équipe Traitement automatique du langage naturel oral	• veille internet • recherche et indexation • text mining • data mining • traduction • traitement de l'image • **traitement de la parole** • représentation graphique • knowledge management	

Adresse :
339, chemin des Meinajaries, Agroparc BP 1228 - 84911 AVIGNON Cedex 9
http://www.lia.univ-avignon.fr/equipes/TALNO/index.html

Effectifs : 21 chercheurs

Axes de recherche et produits :
L'équipe PAROLE du LIA mène des travaux dans le domaine de la reconnaissance et la parole et du locuteur. Six grands thèmes sont abordés :

- Systèmes Embarqués pour le traitement de la parole
- Apprentissage et Adaptation des modèles acoustiques
- Robustesse
- Moteurs de Reconnaissance
- Reconnaissance Multimodale
- Indexation multimédia

Université de Caen – GREYC
Laboratoire GREYC (Groupe de Recherche en Informatique, Image, Automatique et Instrumentation de Caen) UMR 6072
Équipe Données, Document,Langue (DoDoLa)
http://www.greyc.unicaen.fr/

- veille internet
- **recherche et indexation**
- text mining
- data mining
- traduction
- traitement de l'image
- représentation graphique
- knowledge management

Adresse :
Campus Côte de Nacre, boulevard du Maréchal Juin
BP 5186 - 14032 Caen Cedex

Date de création : Le GREYC a été créé en 1995 par la fusion de deux anciens laboratoires.

Effectifs : une quinzaine de chercheurs

Axes de recherche et produits :

Mots-clés : sémantique, extraction d'informations, document électronique, métaphore, sémantique de l'espace, temporalité, référence, modalités, tropes, XML, base de données, document, interactivité, classification, clustering, valeurs manquantes, document mobile, qualité des données, raisonnement à base de cas.

L'équipe a pour ambition de faciliter la production, l'accès et les usages du document électronique mais aussi de façon plus générale aux "données" et ses travaux concernent entre autres le document géographique, la langue et l'extraction de connaissances à partir de bases de données. Les activités scientifiques de l'équipe sont structurées en trois thèmes (brièvement présentés ci-dessous). L'équipe favorise les projets inter-thèmes ainsi que les collaborations avec d'autres équipes et laboratoires. Elle initie et développe des collaborations entre chercheurs issus de domaines et de communautés de pensée différentes autour des multiples facettes recoupant la production, la représentation, la diffusion et l'utilisation de l'information.

- Le groupe *"fouille de données et apprentissage"* propose des modèles interactifs pour améliorer l'appréhension des informations contenues tant dans un document que dans de grandes collections de documents. Il s'intéresse aussi bien aux méthodes (algorithmique de l'extraction) qu'aux multiples usages (e.g., interprétation de données, caractérisation de classes, clustering) des résultats d'un processus d'extraction de connaissances à partir de données.

- Partant du constat que le développement des technologies de l'information et de la communication mène à une situation paradoxale, offrant d'une part de plus en plus de facilités pour la production et la diffusion de documents électroniques, mais rendant d'autre part la recherche et l'accès à l'information de plus en plus complexe, le groupe *"document électronique composite"* étudie la nature et les représentations du document électronique, avec un intérêt particulier pour le document géographique.

- Le groupe *"langue pour le document"* s'intéresse plus spécifiquement à l'information textuelle, élaborant des modèles linguistiques pour la recherche d'information, en articulation avec des travaux sémantiques plus fondamentaux. Les travaux du groupe comportent aussi bien des travaux théoriques visant à proposer des modèles informatiques, notamment en sémantique et en rhétorique et des aspects applicatifs plus directement lié au traitement automatique des langues et permettant de confronter les modèles au matériau linguistique.

Université de Lyon 1 - DOCSI Laboratoire DOCSI (Documents et Sciences de l'information) http://docsi.univ-lyon1.fr/presentation.htm	• veille internet • **recherche et indexation** • text mining • data mining • traduction • traitement de l'image • représentation graphique • knowledge management	

Adresse :
43, bd du 11 Novembre 1918, 69 622 Villeurbanne cedex, Bât Omega

Axes de recherche et produits :

Les recherches de Docsi sont centrées sur le document numérique. Les axes de recherches envisagés se déclinent suivant trois thématiques majeures pour les sciences de l'information :
- Modèles de production et de diffusion du document
- Echanges d'information et structuration des organisations
- Recherche d'information et appropriation des documents.

Université de Montpellier - LIRMM LIRMM (Laboratoire d'Informatique, de Robotique et de Micro-électronique de Montpellier) UMR CNRS 5506 Département Informatique	• veille internet • **recherche et indexation** • **text mining** • data mining • traduction • traitement de l'image • **représentation graphique** • knowledge management	

Adresse :
161 rue Ada - 34392 Montpellier Cedex 5

Effectifs : environ 60 chercheurs

Partenariats industriels privilégiés :
Albert SA

Axes de recherche et produits :

Le département Informatique du LIRMM est structuré en dix équipes dont trois intéressent potentiellement le secteur de la veille :

Ingénierie des Données et des Connaissances (IDC)
Médiation de données dans les systèmes distribués à grande échelle : modèles d'intégration de données axés sur des techniques de fouille de données et maintien de la cohérence.

Traitement algorithmique du langage (TAL)
Analyse morpho-syntaxique et sémantique du Français, classification automatique de documents, recherche d'information et traduction automatique.

Visualisation et algorithme des graphes (VAG)
L'objectif du projet, axé sur la visualisation de graphes, est d'utiliser à la fois des techniques issus de l'algorithmique et la théorie des graphes ainsi que d'outils de visualisation pour des applications telles que le "clustering".

Université de Nantes - LINA Laboratoire d'Informatique de Nantes-Atlantique (LINA) FRE 2729 http://www.sciences.univ-nantes.fr/info/perso/permanents/daille/	• veille internet • **recherche et indexation** • **text mining** • data mining • traduction • traitement de l'image • représentation graphique • knowledge management	

Adresse :
2, rue de la HoussinièreBP 92 20844322 Nantes Cedex 3

Axes de recherche et produits :
Axes : Analyse syntaxique, formalismes grammaticaux, traduction automatique, Fouille terminologique.

Réalisation du logiciel ACABIT : Automatic Corpus-based Acquisition of Binary Terms. ACABIT est un programme d'acquition de terminologie qui prend en entrée un texte annoté linguistiquement et retourne une liste ordonnée de candidats termes.

Université Paris-Sorbonne – LaLIC Equipe Langage, Logique, Informatique et Cognition (LaLIC) CNRS (UMR 8139) www.lalic.paris4.sorbonne.fr	• veille internet • **recherche et indexation** • **text mining** • data mining • traduction • traitement de l'image • représentation graphique • knowledge management	

Adresse :
96 Boulevard Raspail - 75 006 Paris

Axes de recherche et produits :
Les activités de recherche de l'équipe se situent au confluent de trois groupes de disciplines :

- La linguistique (descriptive et théorique) ; modèles cognitifs du langage ;
- La logique et les mathématiques de la cognition ;
- L'informatique et l'Intelligence Artificielle (représentation des connaissances).

Thèmes :

- Filtrage sémantique et structuration des connaissances à partir des textes,
- Collecte et sélection d'information sur le Web,
- Analyse automatique de l'Arabe.

Parmi les travaux de l'équipe, on peut mentionner les travaux portant sur le résumé automatique qui intéresse particulièrement le processus de veille et d'IE.

Université de Paris 7 - TALANA
Laboratoire LATTICE UMR 8094
Équipe TALANA
http://talana.linguist.jussieu.fr/presentation.html

- veille internet
- **recherche et indexation**
- text mining
- data mining
- traduction
- traitement de l'image
- représentation graphique
- knowledge management

Adresse :
TALANA
Pièce 801, Tour centrale 8ème étage
UFRL, Paris 7 2 place jussieu, 75251 Paris cedex 05

Effectifs : une dizaine de membres

Axes de recherche et produits :
Les activités de recherche de portent sur *la Linguistique Informatique* et plus particulièrement :

- Génération de textes
- Interactions sémantique-syntaxe-prosodie
- Modélisation linguistique et dépendance
- Modélisations sémantiques

Université de Paris 13 – LIPN
LIPN (Laboratoire d'Informatique de Paris Nord) - UMR CNRS 7030
Équipe Représentation des connaissances et langage naturel
http://www-lipn.univ-paris13.fr/RCLN/index_fr.php

- veille internet
- **recherche et indexation**
- text mining
- data mining
- traduction
- traitement de l'image
- représentation graphique
- knowledge management

Adresse :
99, avenue Jean-Baptiste Clément 93430 Villetaneuse

Effectifs : 12 membres permanents

Axes de recherche et produits :
Les connaissances sont généralement transmises par l'intermédiaire du langage, ce qui justifie que la Représentation des Connaissances, enjeu majeur de l'Intelligence Artificielle et des Sciences Cognitives, ait pour objectif de pouvoir traiter les connaissances exprimées en Langage Naturel. De manière complémentaire, la compréhension de texte met en oeuvre des connaissances et des raisonnements qu'il est fondamental de décrire et de modéliser.

L'équipe s'intéresse donc au langage, non pas en tant que système formel de signes, mais pour son pouvoir expressif et à la Représentation des Connaissances, en tant qu'outil mis au service du traitement du Langage Naturel.

Les travaux de l'équipe comportent à la fois des recherches à caractère fondamental et des recherches plus appliquées. L'évolution des thèmes de l'équipe a mené à étudier finement pour les modéliser certains phénomènes interprétatifs (liés au pluriel, au temps, aux normes...), à concevoir des outils destinés à faciliter l'accès au contenu de documents techniques ou de vastes bases textuelles, comme le Web, et enfin à développer des outils d'ingénierie des connaissances à partir de textes.

Thèmes de recherche :
- Données, représentations et calculs sémantiques

L'équipe étudie les phénomènes sémantiques d'une façon « fine » (par opposition avec les analyses nécessairement plus simples requises pour un traitement de corpus). En ce qui concerne la sémantique lexicale, leur approche s'oppose à une démarche énumérative qui suppose qu'un lexique peut fournir explicitement la liste des sens des unités, et que le traitement sémantique se limite à trouver l'élément approprié de cette liste. Au contraire, les travaux en cours tentent de rendre compte d'une **dynamique du processus interprétatif** permettant d'adapter le sens des unités lexicales suivant le contexte dans lequel elles apparaissent. Deux types de travaux complémentaires sont menés : d'une part, une étude précise de certaines unités lexicales, d'autre part, une étude sur les procédures interprétatives permettant d'établir la « bonne » valeur sémantique de ces unités en contexte et d'aboutir à des inférences adéquates.

- Sémantique de corpus

L'apport concerne essentiellement l'acquisition des ressources lexicales et conceptuelles auxquelles ces techniques font appel. La construction de ces ressources constitue en effet le goulot d'étranglement des systèmes de traitement de corpus : les dictionnaires de langue sont généralement peu adaptés, les ressources lexicales spécialisées sont rarement disponibles et

toujours difficiles à réutiliser. Nous développons des méthodes et des outils d'aide à la construction de ces ressources lexicales (terminologies, classes sémantiques et schémas prédicatifs) pour un corpus et en fonction d'une tâche donnée. L'élaboration de ces ressources repose sur l'analyse sémantique de corpus de textes.

- Ingénierie des connaissances à partir de textes

Ce thème porte sur l'aide à la construction d'**ontologies** à partir de textes, aide fondée sur une analyse de corpus utilisant des principes linguistiques et des logiciels de traitement automatique de la langue. À la suite de contrats industriels sur la supervision en télécommunications et la détection d'anomalies dans les spécifications informelles en génie logiciel, une méthode de construction d'ontologie a été élaborée parallèlement au développement d'une plate-forme de construction d'ontologie, TERMINAE. Le terme « ontologie » est à prendre au sens large de ressource sémantique, c'est-à-dire index, glossaire, thesaurus, terminologie, réseau conceptuel ou ontologie formelle, allant du moins formel au plus formel.

La version actuelle de TERMINAE est adaptée à la construction de ressources terminologiques. Elle intègre des résultats de logiciels de TAL (Lexter et Syntex de Didier Bourigault de l'ERSS de Toulouse) et en permet le dépouillement. TERMINAE comporte également un concordancier (Linguae) ce qui autorise l'étude des textes d'un point de vue lexical et syntaxique.

Les produits construits dans TERMINAE peuvent être un ensemble de fiches terminologiques, un réseau conceptuel ou une ontologie, constitués de concepts issus des fiches terminologiques et des relations sémantiques entre eux. Ces produits sont facilement exploitables par d'autres (humains ou outils) car toutes les sauvegardes sont en XML. De plus la traçabilité totale des textes aux termes puis aux concepts terminologiques assure la lisibilité des produits élaborés. La version actuelle de Terminae est disponible en français et en anglais.

Une application particulière concerne l'indexation sémantique de documents XML, dans le cadre de corpus spécialisés constitués d'un ensemble de documents partageant une même structure (DTD) et plus précisément sur un ensemble de comptes-rendus d'hospitalisation.

Université de Paris 13 - LLI
LLI (Laboratoire de Linguistique Informatique)
UMR CNRS/Paris 13 7546
http://www-lli.univ-paris13.fr

- veille internet
- recherche et indexation
- **text mining**
- data mining
- traduction
- traitement de l'image
- représentation graphique
- knowledge management

Adresse :
99, avenue Jean-Baptiste Clément
93430 Villetaneuse

Date de création : 1993
Effectifs : 25 membres permanents

Partenariats industriels privilégiés :
SYSTRAN SA

Axes de recherche et produits :
L'objectif fondamental du laboratoire est de décrire le vocabulaire français sur la base de classes sémantiques (technique des classes d'objets).

La définition des mots se fait sur la base de leur environnement, l'unité minimale étant la phrase simple. Le projet consiste à décrire l'ensemble des phrases simples du français en précisant pour chaque emploi prédicatif son schéma d'arguments, les positions argumentales étant définies par les classes d'objets. Un prédicat a autant d'emplois qu'il aura de schémas d'arguments différents. Les classes sémantiques ainsi dégagées reposent donc sur une base syntaxique. Ces classes peuvent être décrites en extension, ce qui permet la génération de toutes les phrases possibles autour d'un emploi prédicatif donné. Sur ces schémas d'arguments se greffe l'actualisation de la phrase, ce qui revient, pour les prédicats, à spécifier leur conjugaison et leur aspect. L'actualisation des prédicats nominaux est prise en charge par les verbes supports, qui font l'objet d'une description systématique.

Les résultats de cette recherche sont intégrés notamment dans les systèmes de traduction automatique.

Université de Provence - Aix Marseillle III - DELIC
Equipe DELIC (DEscription Linguistique Informatisée sur Corpus)
EA 3779
http://www.up.univ-mrs.fr/delic/

- veille internet
- recherche et indexation
- **text mining**
- data mining
- traduction
- traitement de l'image
- représentation graphique
- knowledge management

Adresse :
29, Avenue Robert Schuman 13621 Aix-en-Provence Cedex 1

Effectifs : une dizaine de chercheurs permanents

Axes de recherche et produits :

L'équipe DELIC s'intéresse à la description des structures morphosyntaxiques et lexicales en français, en synchronie comme en diachronie. Sa méthodologie repose sur l'utilisation systématique de grands corpus oraux et écrits, à l'aide d'outils informatiques appropriés (concordanciers, étiqueteurs, outils de gestion et de navigation, etc.).
Les thèmes de recherche sont :

- Construction de corpus
- Alignement de corpus multilingues
- Evaluation d'outils d'alignement (projet Technolangue ARCADE)
- Participation à la campagne d'évaluation EASY (analyseurs syntaxiques).

Université de Provence LPL (Laboratoire Parole et Langage) – UMR CNRS 6057 http://www.lpl.univ-aix.fr	• veille internet • recherche et indexation • **text mining** • data mining • traduction • traitement de l'image • **traitement de la parole** • représentation graphique • knowledge management	

Adresse :
29, avenue Robert Schuman – 13621 Aix en Provence Cedex 1

Effectifs : 26 chercheurs

Axes de recherche et produits :
Les équipes du LPL sont au nombre de six :
1. Appropriation des langues et dysfonctionnements langagiers
2. Production et perception de la parole
3. Analyse des fonctionnements langagiers
4. Créoles
5. Psycholinguistique
6. Prosodie et Représentation Formelle du Langage

Deux équipes intéressent plus particulièrement le secteur de la veille et de l'IE : les équipes 2 et surtout 6 qui travaille dans le domaine du TALN, notamment en syntaxe.

Université de Toulouse II - Le Mirail ERSS (Équipe de recherche en syntaxe et sémantique) UMR 5610 http://www.univ-tlse2.fr/erss/	• veille internet • **recherche et indexation** • **text mining** • data mining • traduction • traitement de l'image • représentation graphique • knowledge management	

Adresse :
5, allées Antonio Machado – 31058 Toulouse Cedex 9

Date de création : 1981

Effectifs : une trentaine de chercheurs

l'ERSS se donne pour fin la description scientifique des langues dans leurs différentes composantes (phonologie, morphologie, syntaxe, sémantique, pragmatique, lexique) et la modélisation des descriptions obtenues. Cette activité donne lieu à des collaborations tant avec les informaticiens (spécialistes de l'intelligence artificielle et de l'ingénierie linguistique) qu'avec les psycholinguistes.

Les langues étudiées sont multiples : au français commun — auquel est consacrée la majorité des travaux de l'équipe —, au latin, à l'anglais, à l'espagnol, au coréen et au japonais, sont venus s'ajouter par exemple au cours des quatre dernières années l'arabe et l'amharique, le barasana et le tatuyo, le sarde, l'italien et le serbo-croate.

Axes de recherche :

- Phonologie : corpus, variation, universaux (Resp. L. Labrune et L.M. Tarrier)
- Morphologie (Resp. N. Hathout et F. Montermini)
- Description syntaxique (Resp. Injoo Choi-Jonin)
- Sémantique et discours (Resp. A. Le Draoulec et J. Busquets)
- Sémantique et Corpus (Resp. A. Condamines) :
- relations conceptuelles en corpus
- variation et genres textuels
- traitements automatiques de corpus
- représentation des connaissances
- Linguistique et dialectologie occitane et romane (Resp. L. Molinu et L. Rabassa)

Université de Savoie - Condillac Équipe Condillac http://ontology.univ-savoie.fr/main.asp	• veille internet • **recherche et indexation** • text mining • data mining • traduction • traitement de l'image • représentation graphique • **knowledge management**	

Adresse :
Université de Savoie - Campus Scientifique
73 376 Le Bourget du Lac cedex - France

Axes de recherche et produits :
Les travaux de l'équipe Condillac portent sur la conception d'ontologies et la visualisation graphique des connaissances, la cartographie sémantique, l'analyse terminologique et sémantique.

Parmi les réalisations de l'équipe : la **OK Sation**.
OCW, pour **Ontology Craft Workbench,** est un environnement logiciel dédié à l'acquisition, la définition et la manipulation de bases de connaissances composées principalement d'ontologies et de réseaux sémantiques. Bien que faisant suite à la **OK Station** (Ontological **K**nowledge **S**tation) **OCW** repose sur des fondements théoriques et pratiques différents.
Les fonctionnalités de **OCW** se présentent sous la forme de modules que chaque **utilisateur** exploite dans son propre environnement (notamment le module linguistique, le module ontologique, et le module des réseaux sémantiques).

Pour toute information, merci de contacter :

Ludovic ETIENNE

Chargé de mission

ludovic.etienne@cigref.fr