

# **Communauté Ingénierie des Connaissances du pôle de compétitivité Cap Digital**

2010



A propos de Cap Digital .....	5
A propos de la Communauté Ingénierie des Connaissances .....	6

## Les projets :

AFI-Gestion 2.0 .....	16
AMAFI .....	24
Angel Studio .....	20
ASAP .....	22
AVEIR .....	9
CADI .....	13
CAHORS .....	18
CatEase .....	25
CEDRES .....	26
Ciné Cast .....	25
CRoTAL .....	13
DECODA .....	22
DaFOE4 App .....	9
Dem@t Factory .....	17
DISCO .....	13
DoXa .....	17
E-fraud Box .....	23
ETL Grid .....	10
Extended Brain .....	26
Feed-ID .....	26
GENIUS .....	14
HUBeNET .....	10
Infomagic 1&2 .....	8
Infomagic 3 .....	12
Jamespills .....	21
LAMPADA .....	23
LegiLocal .....	27
LIMED .....	24
L3IM (Langage Iconique et Interfaces Interactives en Médecine) .....	15
MADSPAM 2.0 .....	14
Media Box .....	22
Medi@Tic .....	11
METRICC (Mémoire de Traduction, Recherche, d'Information et Corpus Comparables) .....	18
NatImages .....	16
Nautilus .....	12
Navigation 3 .....	11

OpenSem .....	20
Open Way III.....	24
RAFFUT .....	14
ROSES .....	15
Samar .....	21
SCARFACE (Caractérisation Sémantique et Recherche de Visages) .....	19
Septia .....	10
SIRE .....	20
Smart Token .....	11
SOPOS (Service Optimisation on Point of Sale) .....	27
TANGUY (Des Textes aux Arguments par Navigation Graphique de l'Utilisateur en Interaction).....	19
Text Mining .....	11
VigiTermes .....	15
VoxFactory.....	18
WEBCROSS .....	15

Cap Digital est le pôle de compétitivité de la filière des contenus et services numériques. Il regroupe plus de 600 adhérents : 530 PME, 20 grands groupes, 50 universités et grandes écoles regroupant 170 laboratoires de recherche.

Le pôle couvre 9 communautés de domaine :

- Culture, Presse, Media
- Design numérique
- Education et Formation Numériques
- Image, Son et Interactivité
- Ingénierie des connaissances
- Jeu Vidéo
- Logiciel libre, Coopération et nouveaux modèles (CoLLibri)
- Robotique et Objets Communicants (Cap Robotique)
- Services et Usages Mobiles

Pour soutenir la créativité et la compétitivité de ce secteur industriel qui représente à lui seul un marché mondial de 300 milliards d'euros, Cap Digital mène les actions suivantes : l'aide au développement de la R&D et de l'innovation ; le développement de plates-formes mutualisées ; l'accompagnement dans le développement des entreprises à tous les niveaux ; la gestion prévisionnelle des compétences et l'accompagnement dans le recrutement, la formation et la gestion de projet collaboratif ; l'organisation d'ateliers et de formations ; la veille et la prospective, notamment à travers son programme « Think Digital » ; et le rayonnement et la compétitivité à l'international.

Depuis 2006, Cap Digital a reçu plus de 1000 projets, en a labellisé plus de 400 parmi lesquels près de 300 ont été financés. Ces projets représentent un investissement total de 550M€, dont environ 250M€ de financement public. En 2009, dans le cadre du plan filière régional, plus de 60 sociétés ont bénéficié d'ateliers d'aide au développement et plus de 40 sociétés ont participé aux missions internationales pilotées par Cap Digital.

[info@capdigital.com](mailto:info@capdigital.com) – [www.capdigital.com](http://www.capdigital.com)

**La communauté Ingénierie des Connaissances rassemble une centaine d'adhérents au sein de Cap Digital. 52 projets collaboratifs de R&D sont menés dans cette thématique. L'investissement total sur ces projets est de 151 M€, dont 65 M€ de financement public.**

La quantité d'informations disponible mondialement double chaque année. Cette prolifération des données, disponibles sur différents média publics ou privés (TV, Internet, bases de données textuelles, enregistrements vidéos, enregistrements sonores...), combinée à leur hétérogénéité (texte, voix, image et vidéo, données numériques) rend la compréhension et l'analyse de l'information de plus en plus difficiles et pourtant indispensables. La croissance exponentielle des informations, loin de rendre l'environnement plus intelligible, en éloigne le sens.

La Communauté Ingénierie des Connaissances regroupe des laboratoires, des PME et l'ensemble de l'écosystème du traitement de l'information et de la « business intelligence » autour de thèmes comme l'extraction sémantique, le traitement de données, leur tri et leur analyse.

Les membres de Cap Digital (entreprises, institutions et laboratoires de recherche) dont les activités sont liées à la Communauté Ingénierie des Connaissances sont :

A2IA	Innovimax	Pass Technologie
Adminext	INRIA	Perfect Memory
Agence Française Informatique	Institut Jean Nicod	PERTIMM
AMI Software	Institut TELECOM	Proxem
Antidot	Intelligence Technologies	Publiva
Ardans sas	INTUIGO	Quatrx
ARISEM	ISTHMA	Quilca
Axalot	JAMESPOT	Qwam Content Intelligence
Beetween	Kadeal	Scan & Target SAS
Bertin Technologiess	KIOLIS	SeAL Online Solutions
Blu-e	KXEN	SOFTISSIMO
Bull	LexisNexis	SQUID SOLUTIONS
CEA	Lillo	SWETS
CLIME	Lingua et machina	Syllabs
Climpact	LINGWAY	TALEND
Crimere	LIP6 - UMR 7606	Technicolor
DistanceWatch	LIPN - UMR 7030	Tecknowmetrix
EADS	Localeo SAS	TEMIS SA
ELLIPSA	Magillem Design Services	Testntrust
ENSTA	MFG R&D	THALES Land & Joint Systems
Ensuite Informatique	MICROSOFT France	TYMIS
ENTROPYSOFT	MIRATECH	Université de Paris 13
Epiphyte	MONDECA	Université de Paris 6
	Moodstock	Université de Paris 5 Université

EPTICA	Nomalys	Paris Dauphine
Exalead	NOMIA	Vecsys
Fist Sa	NOURY SOLUTIONS	Viavoo
Global Imaging Online	NUMSIGHT	Vinci Consulting
Grimmersoft	OWI Technologies	WHP
Hermes Softlab France		XEROX
Hypios		

# Les projets Ingénierie des Connaissances

**La Communauté Ingénierie des Connaissances compte 52 projets innovants pour un budget total de 151 millions d'euros, dont 65 millions d'euros de financement public.**

Abréviations utilisées

*ANR : Agence Nationale de la Recherche*

*W2 : Appel à projets Web 2.0*

*FUI : Fond Unique Interministériel*

## FUI 1 – 1<sup>er</sup> semestre 2006

### Infomagic 1&2

Infom@gic est un projet qui relève de l'ingénierie des connaissances. Ce domaine est celui des Nouvelles Technologies d'Analyse de l'information (NTAI), technologies permettant d'exploiter (recherche, extraction, fusion) de grandes masses d'informations multimédias dont les volumes ne cessent de croître.

Infom@gic vise à mettre en place un laboratoire industriel de sélection, de tests, d'intégration et de validation d'applications intégrant les meilleures technologies franciliennes. Ce laboratoire s'appuie sur une plateforme commune d'interopérabilité couvrant les 3 axes technologiques fondamentaux du domaine : les moteurs de recherche avancés, l'extraction de connaissances, et la fusion d'informations et ce, sur tous les types de sources : données structurées, texte, images & vidéo, et son. La plateforme Infom@gic intégrera des applications développées notamment pour les domaines de l'e-Education, la gestion des patrimoines culturels numériques, la gestion de la relation client, la gestion des risques, la veille, etc.

Résultats attendus : une optimisation de la qualité des résultats produits grâce à l'amélioration des méthodologies et technologies mises en œuvre, des sauts technologiques dans les performances (volumes traités, amélioration de la pertinence des moteurs de recherche), l'interopérabilité des technologies, en particulier pour exploiter simultanément toutes les modalités de l'information ; une plateforme technologique mutualisée au bénéfice de tous les acteurs du domaine ; des démonstrateurs applicatifs qui pourront être industrialisés en vue d'être valorisés commercialement.

**Porteur** : Thalès Communications

**Partenaires Infom@gic 1** :

- Grands groupes : Thalès Communications, EADS, EDF, Editions Odile Jacob
- PME : Bertin Technologies, Xerox, Pertimm, Sinequa, Temis, Vecsys, Fist

- Etablissements publics et partenaires académiques : Académiques : CEA, CNRS, EADS, EDF, INA, INRIA, Institut Telecom, ONERA, Université de Marne La Vallée, Université de Paris 13, Université de Paris 6, Université de Paris 8

#### **Partenaires Infom@gic 2 :**

- Grands groupes : Thalès Communications, EADS, EDF
- PME : Bertin Technologies, Fist, Xerox, Maxicours, Pertimm, Sinequa, Temis, Vecsys Hi, Stor
- Etablissements publics et partenaires académiques : ONERA, Université de Marne La Vallée, Université de Paris 13, Université de Paris 6, Université de Paris 8, CEA, CNRS - INA, INRIA, Institut Telecom

## ANR 2006 – 1<sup>er</sup> semestre 2006

### **AVEIR**

---

L'objectif général d'AVEIR est d'enrichir les systèmes de requêtes par une indexation et une annotation sémantiques et par une description relationnelle symbolique, les unes et les autres bâties à partir d'un traitement simultané des textes associés à l'image et de l'image elle-même telle qu'on les obtient sur le web. Cette information sémantique et symbolique sera utilisée pour réduire l'ambiguïté perceptuelle des images et pour accroître les performances des systèmes de requêtes dans les grandes bases de données.

**Porteur :** Université de Paris 6 – Lip6

#### **Partenaires :**

- Académiques : LIG Grenoble, LSIS Université d'Aix-Marseille, LTCI / Telecom Paris

### **DaFOE4 App**

---

L'objectif du projet est de proposer une méthode complète associée à une plateforme technique pour concevoir des ontologies, de la modélisation à partir du domaine à leur évolution en passant par leur formalisation et exploitation. Il a pour but de prendre en charge la modélisation sémantique des concepts ontologiques (différentiels et formels) pour motiver et justifier les représentations formelles qui seront utilisées et en faciliter la révision. La plateforme technique (DaFOE) est un ensemble d'outils dont un éditeur d'ontologies qui prend en charge toute la question de la sémantique de ces ontologies, à travers des questions épistémologiques liées aux concepts formels de haut niveau et vis-à-vis de la composante métier, à travers des travaux sur les corpus textuels. Le but est de rendre disponible, à travers la plateforme, les méthodologies qui vont permettre de construire l'ontologie.

**Porteur :** INSERM

#### **Partenaires :**

- PME : Mondeca
- Académiques : ENST, Institut Télécom-ParisTech, IRIT, Université de Paris 13-LIPN, LISI, Supélec, UTC / Heudiasyc

## Septia

---

Le projet SEPTIA vise à produire le prototype d'un composant logiciel avancé de traitement des interactions clients dans le cadre d'un centre de contacts. Axé de façon prédominante sur le traitement des emails, le projet étudiera également l'apport des données clients, des données d'interactions inter-clients dans un portail collaboratif.

**Porteur :** Temis

**Partenaires :**

- PME : Eptica, KXEN
- Académiques : Université de Paris 6-Lip6, Université de Paris 13 - LIPN

## CAP DIGITAL 2006 – 1<sup>er</sup> semestre 2006

## HUBeNET

---

TrustSeed SARL, TrustMission SA et Europatrust SA ont créé des activités très innovantes dans le domaine des échanges de documents professionnels à « valeur juridique probante ». L'un des intérêts de la solution dénommée « HUBeNET » c'est d'éviter de rematérialiser les documents informatiques avec une signature manuscrite, et d'authentifier à distance les personnes avec une « identité forte » et avec leurs droits tout en respectant les procédures de gestion documentaire et collaborative (rédaction, signature, expédition, réception, archivage, téléchargement, consultation...).

**Porteur :** Trust Mission

## OSEO – MONOPARTENAIRE – 2<sup>ème</sup> semestre 2006

## ETL Grid

---

Talend est une solution d'intégration de données complète permettant aux entreprises de gérer efficacement leurs données, de les transformer en information métier et d'en extraire toute leur valeur.

**Porteur :** Talend S.A

## Navigation 3

---

Le projet Navigation 3 porte sur le développement et la validation fonctionnelle d'un pilote logiciel pour le marché du tourisme permettant une navigation spatiale, thématique et temporelle dans des environnements 3D multimedia.

**Porteur** : Navidis

## Smart Token

---

Le « token » est le conteneur de métadonnées qui pourra à terme agréger toute la dimension métier des applicatifs. C'est ce « token », qui, externalisé de la Plateforme OpenCS et enrichi en fonctionnalités avancées, va devenir le « smart token », accessible non seulement aux intégrateurs et aux administrateurs mais aussi directement aux utilisateurs finaux.

**Porteur** : Open CS

## Text Mining

---

La plateforme TMS permet à chaque entreprise cliente de développer elle-même ses applications de gestion de la connaissance, sans prendre en compte les besoins et les développements des autres départements.

**Porteur** : Temis

# FUI 4 – 1<sup>er</sup> semestre 2007

## Medi@Tic

---

L'objectif du projet Medi@Tic est de proposer une solution permettant d'enrichir en temps réel, de façon interactive et personnalisée des contenus multimédia reçus par un utilisateur, quel que soit le mode de diffusion. Ce type de service pourrait être proposé par des prestataires qui n'ont pas forcément de lien avec le concepteur ou le fournisseur de contenu. Pour ce faire, MediaTic exploite les méta données déjà disponibles dans le flux diffusé ou en crée automatiquement en temps réel en analysant les contenus. Celles-ci peuvent être aussi collectées à partir de bases de données externes sur Internet, et combinées aux données utilisateur ( profil, contexte, préférences) afin de sélectionner les programmes préférés sur l'ensemble des chaînes disponibles, et de générer des requêtes vers des moteurs de recherche ou des bases de données pour proposer à l'utilisateur de manière interactive des contenus additionnels personnalisés

**Porteur** : Alcatel Lucent

**Partenaires :**

- PME : Expway, La Banque Audiovisuelle (Vodeo), Vecsys, New Phenix
- Grandes Entreprises : Groupe Bull, Radio France, SFR
- Académiques : CEA, Institut TELECOM, Supélec



---

Infom@gic est un projet qui relève de l'ingénierie des connaissances. Ce domaine est celui des Nouvelles Technologies d'Analyse de l'information (NTAI), technologies permettant d'exploiter (recherche, extraction, fusion) de grandes masses d'informations multimédias dont les volumes ne cessent de croître.

Infom@gic vise à mettre en place un laboratoire industriel de sélection, de tests, d'intégration et de validation d'applications intégrant les meilleures technologies franciliennes. Ce laboratoire s'appuie sur une plateforme commune d'interopérabilité couvrant les 3 axes technologiques fondamentaux du domaine : les moteurs de recherche avancés, l'extraction de connaissances, et la fusion d'informations et ce, sur tous les types de sources : données structurées, texte, images & video, et son. La plateforme Infom@gic intégrera des applications développées notamment pour les domaines de l'e-Education, la gestion des patrimoines culturels numériques, la gestion de la relation client, la gestion des risques, la veille, etc.

**Porteur :** Thalès Communications

**Partenaires Infom@gic 3 :**

- Grands groupes : Thalès Communications – EADS, EDF
- PME : Xerox, Maxicours, Pertimm, Sinequa, Temis, Vecsys
- Etablissements publics et partenaires académiques: CNRS, INA, Institut Telecom – ONERA, Université de Marne La Vallée, Université de Paris 13, Université de Paris 6, Université de Paris 8

<b>OSEO – MONOPARTENAIRE – 1er semestre 2007</b>
--



---

Nautilus est une plateforme de services qui a vocation à se connecter à un ensemble d'outils et de procédures qui constituent l'environnement de création de connaissance client d'un opérateur télécom. Cette approche privilégie une démarche dont les concepts ont été posés récemment mais qui n'a pas encore de traduction opérationnelle : le marketing exploratoire. Cette activité consiste à permettre au département marketing de procéder à des recherches de nouvelles formes d'éléments de connaissance client dans les données.

**Porteur :** Squid Solutions

## ANR 2007 – 1<sup>er</sup> semestre 2007

### CADI

Le projet CADI vise à produire un prototype de composants logiciels pour la Distribution, tant brick-and-mortar (magasins) que vente en ligne (site de e-commerce). Il met en œuvre des techniques de data mining et réseaux sociaux intégrées dans des composants logiciels et validées à travers deux démonstrateurs en vraie grandeur dans un magasin et sur un site de e-commerce de biens immatériels. Il permet de fournir des recommandations au consommateur au moment de l'achat, permettant d'augmenter la fidélité client et le volume des ventes.

**Porteur :** Kxen

**Partenaires :**

- Grands groupes : Groupe SAMSE
- PME : Mondomix, Numisight, Spacecode
- Académiques : Université de Paris 6 – LIP6, Université de Paris 13 – LIPN, INSA de Rouen – LITIS

### CRoTAL

Le projet CRoTAL (Conditional Random Fields pour le TAL) a pour objectif d'étudier et de développer de nouvelles techniques pour la manipulation de grandes masses de données textuelles, en se concentrant plus particulièrement sur une technique d'apprentissage automatique particulièrement innovante : les « Conditional Random Fields » ou « champs conditionnels aléatoires ». Il s'agit d'une famille de modèles graphiques introduits dans le contexte de la linguistique informatique, qui permettent d'annoter des données à partir d'exemples de données déjà annotés : ils se situent actuellement au meilleur niveau dans plusieurs domaines, notamment pour l'extraction d'information et la structuration des données textuelles.

**Porteur :** LIMSI, Université Paris Sud XI

**Partenaires :**

- Académiques : Institut Telecom-ENST, Université de Paris 13, Université d'Orléans-LIFO

### DISCO

Le projet DISCO a pour objectif la conception et l'évaluation de techniques génériques et flexibles de recherche et d'indexation basées sur le contenu, et dédiées à des sources de documents multimédia distribués.

**Porteur :** Université Paris-Dauphine-LAMSADE

**Partenaires :**

- Académiques : CEDRIC–CNAM, IRCAM, LAMSADE

Etablissements publics : Agence photographique de la Réunion des Musées Nationaux

- Associations : European Web Archive

## GENIUS

---

Le résultat attendu de ce projet est un logiciel capable de supporter l'activité des designers en conception, et plus précisément les deux phases que sont la catégorisation d'images et la génération de formes. Ce logiciel est découpé en trois volets : la catégorisation d'images (pour l'extraction de composantes design), la génération de formes (pour la génération de modèle), et la visualisation (interface de réalité virtuelle).

**Porteur** : ENSAM–LPCI

**Partenaires** :

- PME : Virtools, Diedre Design
- Académiques : Université de Paris 6 – Lip6

## MADSPAM 2.0

---

Le projet vise à développer une bibliothèque d'outils génériques permettant de s'adapter aux différentes sources et manifestations du spam, non seulement dans les deux applications cibles du projet (moteurs de recherche, blogs et forums) mais aussi dans les applications proches comme les wiki et des domaines émergents comme les réseaux sociaux. Pour ces différentes applications, l'interactivité et l'adaptabilité seront privilégiés.

**Porteur** : France Télécom R&D– Portail Orange (moteur Voilà)

**Partenaires** :

- PME : BlogSpirit – KartOO
- Académiques : Université de Paris 6 – Lip6

## RAFFUT

---

Le projet RAFFUT (Repérage Automatique de Fichiers Frauduleux sur sites UGC) permet d'aboutir à une plate-forme commerciale de détection, de protection et de gestion de droits d'auteurs et de propriété intellectuelle et industrielle utilisable dans le contexte spécifique des sites UGC.

**Porteur** : Advestigo

**Partenaires** :

- PME : SportFive
- Académiques : INRIA LPES

## ROSES

---

Le projet ROSES (Really Open Simple Efficient Syndication) a pour objectif la définition et l'implantation de nouveaux services de syndication web pour la génération, l'agrégation et la personnalisation de flux RSS. Les solutions proposées sont fondée sur des techniques récentes d'interrogation de flux de données XML

**Porteur** : Université de Paris 6 –LIP6

**Partenaires** :

- PME : 2or3things
- Académiques : Conservatoire National des Arts et Métiers, Université de Versailles Saint-Quentin en Yvelines, Université du Sud Toulon-Var

## VigiTermes

---

Le projet VigiTermes vise à améliorer l'environnement de travail des équipes de pharmacovigilance. Les outils et ressources développés sont intégrés sur une plate-forme de gestion des connaissances couplée à des outils de fouille de texte pour améliorer l'accès, l'analyse et la documentation des cas de pharmacovigilance.

**Porteur** : Université de Saint Etienne – DSPIM

**Partenaires** :

- PME : Mondeca, Temis,
- Académiques : IM&Bio, INALCO, INSERM, Rennes 1 EA3888, LORIA, Saint Etienne – DSPIM
- Centre collaborateur de l'OMS pour la pharmacovigilance, Centre de pharmacovigilance de l'HEGP

## WEBCROSS

---

WEBCROSSLING est un service web 2.0 de traduction automatique basé sur une technologie de recherche interlingue associant les utilisateurs pour l'amélioration des dictionnaires

**Porteur** : CEA-LIST Fontenay

**Partenaires** :

- PME : NewPhenix, Softissimo

<b>ANR 2008 – 1<sup>er</sup> semestre 2008</b>
--

## L3IM (Langage Iconique et Interfaces Interactives en Médecine)

---

Le projet L3IM vise à améliorer l'ergonomie de la consultation des données du dossier médical informatisé en ville ou à l'hôpital, un enjeu médico-économique important qui peut

avoir un retentissement sur la qualité des soins prodigués, sur la productivité des médecins et des équipes soignantes et sur la qualité des échanges d'informations médicales entre médecins et autres acteurs de santé. Il a aussi pour objectif de faciliter l'accès aux connaissances contenues dans les guides de bonnes pratiques cliniques et la consultation des moteurs de recherche médicaux comme CISMeF.

Il permet de développer des approches graphiques basées sur un langage iconique breveté par l'Université Paris 13 et pouvant conduire à des solutions nouvelles pour fournir des réponses à des problèmes d'accès à des informations médicales notamment en situation clinique.

**Porteur** : Université de Paris 13-LIM &Bio

**Partenaires** :

- Grands groupes : Mc Kesson, Vidal
- PME : Silk Informatique, SNR
- Académiques : CISMeF, Evalab, SFTG

## OSEO – MONOPARTENAIRE – 1<sup>er</sup> semestre 2008

### AFI-Gestion 2.0

---

Le projet AFI est celui d'une gamme de logiciels de gestion en FAH (Fournisseur d'Accès Hébergé) à destination des petites structures territoriales reliée à une base de connaissance d'analyses réglementaires collaboratives.

**Porteur** : AFI

## ANR 2008 – 1<sup>er</sup> semestre 2008

### NatImages

---

Le projet NatImages développe de nouvelles méthodes de traitement d'images, adaptées aux données ayant des structures géométriques complexes. Les contours et les textures présentent dans les images naturelles ont une régularité anisotrope qui nécessite l'utilisation de représentations généralisant la transformée de Fourier ou en ondelettes. Ces méthodes géométriques ont des applications importantes en astrophysique, où la séparation des différentes sources nécessite des aprioris optimisés. Une autre classe d'applications sont les neurosciences, et en particulier le traitement de l'imagerie optique corticale.

**Porteur** : Université de Paris 9-CEREMADE

**Partenaires** :

- Académiques : CEA, CMLA, ENS Cachan, GREYC, ENSI Caen

## FUI 6 – 1<sup>er</sup> semestre 2008

### Dem@t Factory

---

Le projet DEM@T FACTORY vise à améliorer ou développer tous les outils et toutes les méthodes propres à mettre en place une chaîne industrielle de numérisation du patrimoine, tous types de documents écrits confondus.

DEM@T FACTORY traite plusieurs thématiques : le développement d'un superviseur «intelligent» pour la gestion des flux et des problèmes, la modélisation des documents permettant de le gérer dans son ensemble et non page par page, des outils de prédiction de performance à tous les niveaux, la visualisation 3D des documents pour faciliter l'intervention humaine et le développement d'interfaces pour augmenter la rapidité de l'adaptation à un nouveau type de documents.

**Porteur :** SAFIG

**Partenaires :**

- PME : A2IA, BancTec, Temis
- Académique : CNAM-CEDRIC, ESIEE-A2SI, Université de Paris 6-LIP6

### DoXa

---

Le projet DoXa adresse un ensemble de problématiques liées au traitement automatique des opinions et sentiments dans des corpus de données multilingues, intégrant à la fois des données non- structurées et structurées

Les chaînes de traitement développées pour les end-users aideront à l'observation dynamique, quantitative et qualitative, du positionnement des consommateurs, clients, utilisateurs, des rapports qu'ils entretiennent avec les univers à propos desquels ils s'expriment ainsi que les tendances ou évolutions à l'œuvre dans ces univers. Elles aideront à améliorer les processus d'analyse décisionnelle (olap, segmentation, scoring, etc.) en intégrant dans ces processus des connaissances enrichies. Les travaux doivent permettre de concrétiser les résultats de la recherche industrielle et académique des partenaires du Consortium à des fins de design de nouveaux services, destinés à être commercialisés. Le projet se situe en ce sens dans une perspective de développement pré-compétitif.

**Porteur :** Thalès

**Partenaires :**

- Grands groupes : ARISEM, EDF R&D/ICAMM, MEETIC, Thalès,
- PME : ELDA (Partenaire associé), IObjects, Opinion Way, Pertimm
- Académiques : GM, CNRS- LIMSI, LIR, Université de Paris6 - Lip6, Université de Paris 8 - CHart & LUTIN

## VoxFactory

---

Le projet Vox Factory propose d'analyser et de modéliser la qualité et l'efficacité de l'interaction client/télé-conseiller en s'appuyant sur les technologies du traitement automatique de la parole, de la détection des émotions et du text mining. L'approche est une approche mixte fondée d'une part sur l'analyse textuelle de l'appel, à partir de sa transcription automatique, et d'autre part sur la prise en compte de paramètres physiques du signal de parole pour la détection des émotions notamment.

**Porteur** : Vocalcom

**Partenaires** :

- Grands groupes : EDF
- PME : Safig, Sinequa, Temis, Vecsys, Vecsys Research
- Académiques : CNRS – LIMSI, LEM

## ANR 2008 – 1<sup>er</sup> semestre 2008

## CAHORS

---

Le projet CAHORS est une plate-forme logicielle pour assurer l'interconnexion des outils d'information pour que l'ensemble de la chaîne de traitement puisse être véritablement maîtrisée. Il propose de relever chacun de ces défis techniques, à la fois dans une optique de protection du citoyen face aux menaces terroristes ou criminelles, et pour assurer que la première étape de la gestion de crise potentielle, celle de l'anticipation et de la détection des signes précurseurs, soit idéalement franchie pour une meilleure prévention des risques.

**Porteur** : Thalès Communications

**Partenaires** :

- Grands groupes : Thalès Research
- PME : Arisem
- Académiques : Institut Jean Nicod, Université de Paris6 – Lip6, ONERA

## METRICC (Mémoire de Traduction, Recherche, d'Information et Corpus Comparables)

---

Le projet METRICC exploite les possibilités offertes par les corpus comparables dans le cadre de trois applications industrielles concrètes : les mémoires de traduction, la recherche d'informations interlingue et la catégorisation multilingue.

**Porteur** : Université de Nantes

**Partenaires :**

- PME : Lingua et Machina, Sinequa, Syllabs, Valoria
- Académiques : Laboratoire d'Informatique de Grenoble, Laboratoire Information de Nantes-Atlantique

## SCARFACE (Caractérisation Sémantique et Recherche de Visages)

---

Le projet vise à réaliser des outils d'aide à la recherche d'individus dans des lieux publics déjà équipés de réseaux de caméras de vidéosurveillance standard (un contexte typique correspond aux infrastructures de la RATP), dans un contexte d'élucidation de délits. L'enquêteur doit pouvoir utiliser des informations provenant de témoignages (témoins, victimes, etc.) ou de désignation visuelle sur une vidéo pour retrouver rapidement des informations dans de grandes bases de données de vidéo-surveillance. Pour mettre en place un tel outil de recherche, des algorithmes de traitement d'images (reconnaissance de formes) sont développés et évalués pour détecter, suivre et caractériser les personnes par leurs traits caractéristiques (notamment du visage).

**Porteur :** Université de Caen – INRIA-LEAR

**Partenaires :**

- Grands groupes : EADS Innovation Works
- PME: SpikeNet Technology
- Académiques : Université de Caen, INRIA-IMEDIA, Université de Lille 2 – IREENAT
- Site web: <https://scarface.greyc.fr>

## TANGUY (Des Textes aux Arguments par Navigation Graphique de l'Utilisateur en Interaction)

---

Le projet TANGUY vise à aider des experts métier à dégager les faits les plus pertinents dans des documents et à rechercher des relations entre ces faits permettant d'étayer une chaîne de raisonnement. Sur le principe d'une découverte par émergence, l'utilisateur explore interactivement sa base de connaissance, requête et visualise ses données, définit ses propres centres d'intérêt. Le projet s'appuie sur l'utilisation d'outils d'extraction d'information, de réseaux sémantiques et de visualisation de données.

**Porteur :** Thales Communications France

**Partenaires :**

- Grands Groupes : Fidal, Xerox
- Académiques : INRIA Bordeaux Sud-Ouest

## OSEO – EUREKA– EUROSTARS – 1<sup>er</sup> semestre 2008

### OpenSem

OpenSem est une plateforme ouverte de services dédiés aux traitements sémantiques des contenus du web.

**Porteur :** Exalead

**Partenaires :**

- PME : Priberam Informatica, Synapse Développement

## FEDER 0 – 2<sup>e</sup> semestre 2008

### SIRE

SIRE se donne pour objectif de développer des technologies permettant automatiquement de récupérer les offres d'emploi disponibles sur Internet, d'en faire des traitements sémantiques permettant d'en lever l'implicite afin de produire des catégories d'analyse. Ce dispositif a vocation à constituer à terme la base technique d'un observatoire du marché du travail, diffusant des informations tant quantitatives que qualitatives sur les offres d'emploi en ligne.

**Porteur :** Lingway

**Partenaires :**

- PME : Proxem
- Académiques : MoDyCo

## FEDER 1 – 2<sup>e</sup> semestre 2008

### Angel Studio

Le projet Angel Studio a pour objectif de développer en 24 mois un générateur d'avatars permettant de produire à partir de photos et d'échantillons de voix– correspondant à un utilisateur donné–un avatar de ce dernier, sous la forme d'un visage 3D photo–réaliste, animé en temps réel et doté d'une voix de synthèse paramétrable.

**Porteur :** As An Angel

**Partenaires :**

- PME : As An Angel, SoBuzzy
- Académiques : IRCAM, Télécom Sud Paris

## FUI 7 – 2<sup>e</sup> semestre 2008

### Samar

Le projet SAMAR a pour objectif de développer une plateforme de traitement multimédia en langue arabe, et de valider ces travaux en menant les expérimentations sur la totalité des dépêches arabes jusqu'ici produites par l'AFP (environ un million de dépêches, représentant plus de 150 millions de mots) ainsi que sur des données radio/télévisées en langue arabe. Cette plateforme intéressera non seulement l'AFP, mais également tous les médias arabes de la bordure Méditerranéenne et du Moyen-Orient.

**Porteur :** Temis

**Partenaires :**

- Grandes groupes : Agence France Presse
- PME: Antidot, Cermon, Mondeca, Nuxeo, Vecsys, Vecsys Research
- Académiques : Inalco, Université de Caen– Greyc, Llacan (CNRS), CNRS–LIMSI

## THD – 1<sup>er</sup> semestre 2009

### Jamespills

Ce projet vise à fournir un service de recherche étendu permettant aux utilisateurs de bénéficier des capacités du très haut débit pour la compréhension/visualisation des résultats. Le service enrichit une liste de documents résultats de Google, pour présenter plusieurs listes et types de résultats, provenant de plusieurs autres moteurs de recherche, basés sur des algorithmes de pertinence automatiques ou sociaux. Jamespills doit devenir une plateforme capable d'accueillir par simple configuration plusieurs moteurs et plusieurs réseaux. Pour le grand public, cela permettra de rechercher un film et d'avoir simultanément les infos sur le film et les commentaires sur des réseaux sociaux. Dans l'entreprise, on pourra proposer de rechercher avec Google, et de retrouver les espaces collaboratifs de l'Intranet qui traitent des résultats.

**Porteur :** Jamespot

**Partenaires :** Consortium THD

- Grands groupes : Orange Labs, Groupement Cartes Bancaires
- PME : Bearstech, FaberNovel
- Académiques : Cité des Sciences et de l'Industrie, Centre Pompidou, Institut Télécom, IRI, MSH Paris Nord, Telecom Paris Tech, Université Paris 13
- Associations : AFNIC

## OSEO – ISI– 1<sup>er</sup> semestre 2009

### Media Box

Le projet Media Box est un moteur de veille pluri-media sur lequel les dirigeants, les hommes politiques, les professionnels de la communication ou du marketing, les cadres d'entreprise mais aussi les salariés ou des particuliers pourront faire des recherches ciblées sur l'ensemble des médias. Il est donc un complément des moteurs généralistes pour tous ceux qui veulent suivre l'ensemble des médias (presse écrite, TV, radios et naturellement Internet et blogs) et obtenir une synthèse qualifiée de l'information par rapport à leurs critères.

**Porteur :** PressIndex

**Partenaires :**

- PME : Antidot, Document Image Solutions, Qwam Content Intelligence, Spigraph, Temis, Vecsys, Vecsys Research

## ANR 2009 – 1<sup>er</sup> semestre 2009

### ASAP

ASAP a pour objectif de créer une rupture dans la théorie et la méthodologie de l'apprentissage statistique, en s'attaquant à l'un de ses verrous majeurs, à savoir la création automatique de caractéristiques pertinentes. Les objectifs scientifiques du projet concernent à la fois des aspects théoriques et algorithmiques de l'apprentissage statistique. Le produit de ces analyses sera une plate-forme logicielle qui sera mise à la disposition de la communauté scientifique et qui servira de base à des démonstrations de cette nouvelle approche sur des tâches d'apprentissage réputées difficiles.

**Porteur :** LITIS – Université de Rouen

**Partenaires :**

- Académiques : GREYC-ENSICAEN, Lif-Université Aix Marseille, LIP6-UPMC Paris6, TAO-INRIA

### DECODA

Le but du projet DECODA est de permettre le développement et la gestion sur le long terme des services de speech analytics (analyse automatique de conversations) à un moindre coût, en permettant d'alléger l'effort d'annotation nécessaire en permettant d'alléger l'effort d'annotation nécessaire. Le projet propose de développer des systèmes de fouille robuste de données audio, dans le cadre applicatif de centres d'appels, en utilisant à la fois des ressources linguistiques génériques et des méthodes d'apprentissage peu supervisé.

**Porteur :** Université d'Avignon

**Partenaires :**

- PME : SoNear
- Grands groupes : RATP
- Académiques : LIF (Laboratoire d'Informatique Fondamentale de Marseille)

## E-fraud Box

---

Le projet E-fraud Box vise à développer une boîte à outils intégrée, dédiée à la détection et à l'investigation de la fraude à la carte bancaire sur Internet.

La boîte à outils permettra de répondre à deux besoins majeurs et complémentaires des utilisateurs finaux :

- Mettre en œuvre des analyses visant à identifier le plus rapidement possible les cartes utilisées frauduleusement sur Internet, ce qui permettra au GIE Cartes Bancaires « CB » de bloquer les demandes d'autorisation
- Modéliser les modes opératoires des fraudeurs connus des forces de sécurité afin d'identifier et détecter rapidement de nouveaux modes opératoires

**Porteur :** Thalès Communications

**Partenaires :**

- Grands groupes : GIE CB, Thalès Communications
- PME : Altic, KXEN
- Académiques : Université de Paris 6 – Lip6, LIPN

## LAMPADA

---

LAMPADA est un projet de recherche fondamentale qui traite du passage à l'échelle des méthodes d'apprentissage automatique dans le contexte du traitement des données structurées. Il vise à étudier et développer des algorithmes pour trois grands défis qui sont, les grandes dimensions, les masses de données et la prise en compte de la dynamique des données. Les données structurées apparaissent dans de très nombreux domaines comme la biologie, le multimédia, la vidéo, le web, les applications web 2.0.

**Porteur :** INRIA

**Partenaires :**

- Académiques : Laboratoire Hubert Curien, Laboratoire d'Informatique Fondamentale de Marseille, Université de Paris 6 – Lip6

## FUI 8 – 1<sup>er</sup> semestre 2009

### Open Way III

---

L'objectif du programme Open Way III est de fluidifier le marché de l'emploi sur Internet en maximisant les opportunités qualifiées de mise en relation entre candidats et offres d'emploi via Internet. Il vise en particulier à faire émerger une deuxième génération d'outils pour le recrutement en ligne.

**Porteur** : Quatuor Solutions

**Partenaires** :

- PME : Exalead
- Académiques : Université de Paris 6 – Lip6
- Associations : AGEFIPH

## FEDER 2– 1<sup>er</sup> semestre 2009

### LIMED

---

LIMED est un projet de moteur Meta-Educatif Linguistique pour Contenus Audiovisuels. Il propose de générer automatiquement des quiz de compréhension sur des contenus audiovisuels authentiques préexistants et fortement attractifs, par exemple des séries TV ou des news en langue originale

**Porteur** : 4N media group

**Partenaires** :

- PME : Jamespot
- Académiques : Institut Telecom

## FUI 8 – 1<sup>er</sup> semestre 2009

### AMAFI

---

AMAFI est un projet de prototype de composants logiciels voué à l'analyse de séries financières massives. Ces données devront être acquises, prétraitées, analysées par des méthodes numériques, d'éconophysiques, de détection de formes, et de prévision de séries temporelles. L'ensemble de ces techniques s'illustrera dans un démonstrateur fonctionnel, destiné aux analystes de BBSP.

**Porteur** : BBSP

**Partenaires :**

- PME : BBSP, KXEN
- Académiques : CNAM, CRSA, Université de Paris 6 – Lip6

 **Ciné Cast**

---

Le projet **CineCast** s'appuie sur une plateforme industrielle de diffusion des films et des métadonnées (infrastructure Globe Cast / Orange et outils d'indexation et de gestion de média de sa filiale NETIA) et développe des outils permettant de nouvelles formes de navigation, d'enrichissement de contenus et de partage de points de vue par le plus grand nombre. CineCast combine pour cela le savoir-faire éditorial et en matière d'ingénierie documentaire des principales bibliothèques de films françaises et du premier média social du cinéma Allo Ciné, des technologies d'analyse automatique de l'image, des outils d'ingénierie sociale (moteurs de recherche, technologies collaboratives) et enfin de nouvelles interfaces d'annotation, de rendu et d'échange stimulant la création de nouveaux services et de start-ups.

**Porteur :** NETIA

**Partenaires :**

- Coordination scientifique et technique : Centre Pompidou-IRI et CNRS-LIRIS
- PME : BPI, Exalead, Jamespot, Globe Cast, Lesite.TV, Univers Ciné, VodKaster
- Académiques: AlloCiné, BNF, Centre Pompidou-IRI, CEA List, Cinémathèque Française Forum des Images, INA, CNRS- Liris, Institut Telecom – ParisTech

<b>W2 2009 – 2ème semestre 2009</b>
-------------------------------------

 **CatEase**

---

CatEase propose une interface 2.0 innovante pour l'affichage et la vente des produits en ligne, permettant une présentation attractive des produits, dynamique en fonction de l'arrivée de nouveaux produits à afficher, personnalisée à l'utilisateur, autorisant un accès par critères d'indexation multiples allant au-delà d'une structure arborescente classique et une visualisation simultanée des contenus de plusieurs produits. Elle est enrichie d'un système de personnalisation originale et de possibilités d'actualisation automatique. Cette interface, conçue de façon générale, est mise en œuvre pour un catalogue de librairie numérique avec ePagine et testée par des représentants des trois maillons de la chaîne du livre (éditeurs, libraires et bibliothécaires).

**Porteur :** iObjects

**Partenaires :**

- PME : ePagine
- Académiques : CHart, Université de Paris 6 – Lip6

## CEDRES

---

Le projet CEDRES vise à produire un ensemble de composants permettant d'exploiter les données des sites sociaux dans un modèle économique viable. Le projet est très novateur aussi bien du point de vue scientifique (capacité à développer des algorithmes de data mining « sociaux ») que du point de vue technique (capacité à déployer des technologies innovantes dans des environnements temps réel, à forts volumes de données). Les cinq PME membres du Consortium attendent du projet des retombées très concrètes pour leur activité.

**Porteur :** KXEN

**Partenaires :**

- PME : AF83, Heaven, Telefun – Skyrock,
- Académiques: Université de Paris 6 – Lip6, Université de Paris13 – L2TI

## Extended Brain

---

Extended Brain est une application de traitement participatif de la documentation numérique. La profusion documentaire apportée par le Web est aujourd'hui plutôt un problème qu'une solution, en particulier pour le jeunes utilisateur en situation explicite ou implicite d'apprentissage : télécharger, bookmarker, s'abonner sont des actions qui offrent de nouvelles ressources

Extended Brain appartient à une nouvelle génération d'éditeurs de contenus, tournés vers l'appropriation, l'agencement, la création, la coopération. L'interface transforme la façon de prélever des ressources, de les qualifier extrêmement rapidement, grâce à un moteur sémantique; il les organise, quels que soient leur source et leur format, en une page de travail, ouverte à la participation d'une communauté impliquée.

**Porteur :** PlasmaSoft

**Partenaires :**

- PME : Les Ateliers, Proxem
- Académiques : Ensci

## Feed-ID

---

La plateforme Feed-ID propose une base de sources qualifiées à l'aide de méthodes automatiques ou semi-automatiques. Le projet mettra cette base ouverte à disposition avec les APIs permettant de l'interroger. En parallèle, les fonctionnalités du traitement du contenu textuel utilisé seront également mises à disposition.

Les verrous concernent l'extraction de contenu d'une page web, l'expansion de la base en découvrant automatiquement de nouvelles sources dans un thème donné, la fusion de méthodes topologiques et « sémantiques » et l'adoption en termes d'usage de la plateforme.

**Porteur :** Syllabs

**Partenaires :**

- PME : RTGI, Syllabs, Wikio

## FUI 9 – 2ème semestre 2009

### LegiLocal

---

Le projet LegiLocal vise à rendre l'administratif et juridique accessible au citoyen. Il vise à simplifier l'accès des citoyens aux données administratives, juridiques, publiques locales en proposant aux collectivités une solution qui s'intègre aisément à leur site web en s'appuyant sur les technologies du Web Sémantique. Techniquement, il s'agit de faciliter l'accès au droit via des web services simples d'interrogation et de consultation et de mettre à disposition des widgets d'exploitation de ces web services dans les interfaces utilisateurs dédiées.

**Porteur :** LexisNexis

**Partenaires :**

- PME : Jamespot, LexisNexis, Mondeca, Temis
- Académiques : CERSA, LIPN

## FEDER 3 – 2ème semestre 2009

### SOPOS (Service Optimisation on Point of Sale)

---

Le projet SOPOS a pour ambition de mesurer et d'optimiser la qualité de service apportée à un consommateur lors de son parcours en magasin à travers les deux indicateurs suivants : la qualité de l'accueil lors de son entrée en magasin et son temps d'attente en caisse.

Le projet consiste à analyser en temps réel des flux vidéo en provenance de caméras situées en magasin et en caisse à l'aide d'algorithmes de traitement d'images. Ensuite, il s'agira d'interpréter les mesures observées et d'en déduire des comportements humains. Enfin, ces données seront intégrées à un outil expert d'allocation de forces de vente dans le but de fournir des indicateurs de qualité de service aux responsables de magasins.

**Porteur :** CLIRIS

**Partenaires :**

- PME : Holy – Dis, IBISC



info@capdigital.com  
Tél. +33 (0)1 40 41 11 60  
Fax. +33 (0)1 40 41 11 65  
Passage de l'Innovation  
74 rue du Faubourg Saint-Antoine – 75012 Paris

Notre actualité : [www.capdigital.com](http://www.capdigital.com)  
Nos événements : [www.digitallyours.fr](http://www.digitallyours.fr)  
[www.twitter.com/Cap\\_Digital](https://www.twitter.com/Cap_Digital)  
[www.facebook.com/capdigital](https://www.facebook.com/capdigital)