

Reveal-This

Ou comment générer des
métadonnées utiles
automatiquement



Consortium

Académiques



Développement



Développement



Privés

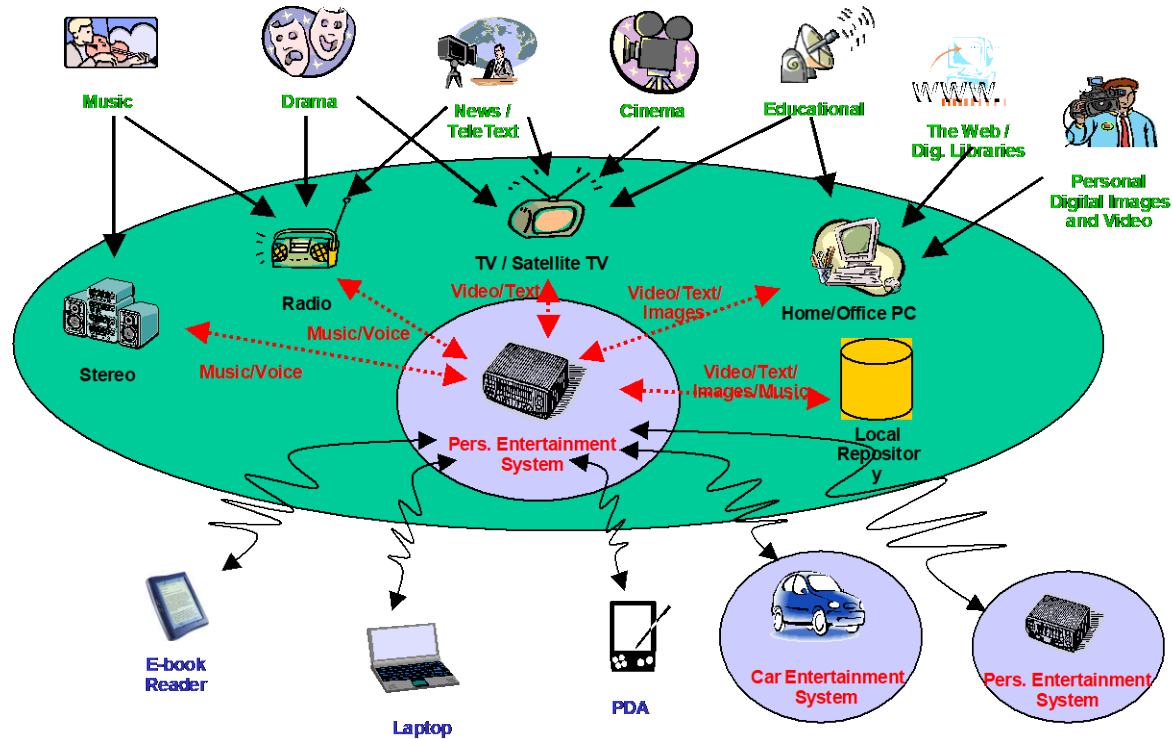
Utilisateurs



Objectifs

- Sources d'informations multiples
 - Papier
 - Web
 - Télévision
 - Radio
- Recherche de l'information de plus en plus compliquée
 - Services à la demande pertinents

Description générale



Généralités

- Améliorer le contenu de documents multimédia par de l'information sémantique:
 - Vignettes significatives
 - Locuteurs identifiés
 - Visages
 - Nom propres
 - Informations sur les sujets et les faits
 - Hyperliens automatiques

Méthodes

- Suite de CIMWOS
 - Utilisation de
 - Parole
 - Reconnaissance et transcription de la parole
 - Reconnaissance du locuteur
 - Image (vidéo)
 - Détection de formes (visages)
 - Reconnaissance de visages
 - Texte
 - Reconnaissance des textes

Eléments de recherche (IST)

- ASR (Automatic Speech recognition)
 - Moteur destiné à reconnaître le locuteur et assurer la transcription vers le texte
- La détection de forme et plus particulièrement le visage ainsi que sa reconnaissance
- Traitement des éléments textuels

Éléments de recherche

- Analyse de texte
 - Catalogage automatique
 - Base d'indexation
 - Résumé automatique
 - Multilinguisme (traduction automatique)
 - Représentations multilingues
 - Éléments de recherche basés sur les mêmes outils
 - Profils

Indexation

- Indexation croisée entre médias
- Indexation croisée entre moteurs d'analyse
- Classification thématique

Applications

- Contenu poussé
 - Sur base d'un profil
 - Dynamique
 - Statique
 - Sur base de demandes ciblées ou précontraintes (mots clés)
 - Par approches successives (intuitives)

Applications

- Contenus tirés
 - Requêtes complexes
 - Professionnels
 - Logiques floues
 - Pertinence des réponses
 - Niveau de confiance dans la réponse
 - Qualité des réponses obtenues
 - Approches successives à mémoire

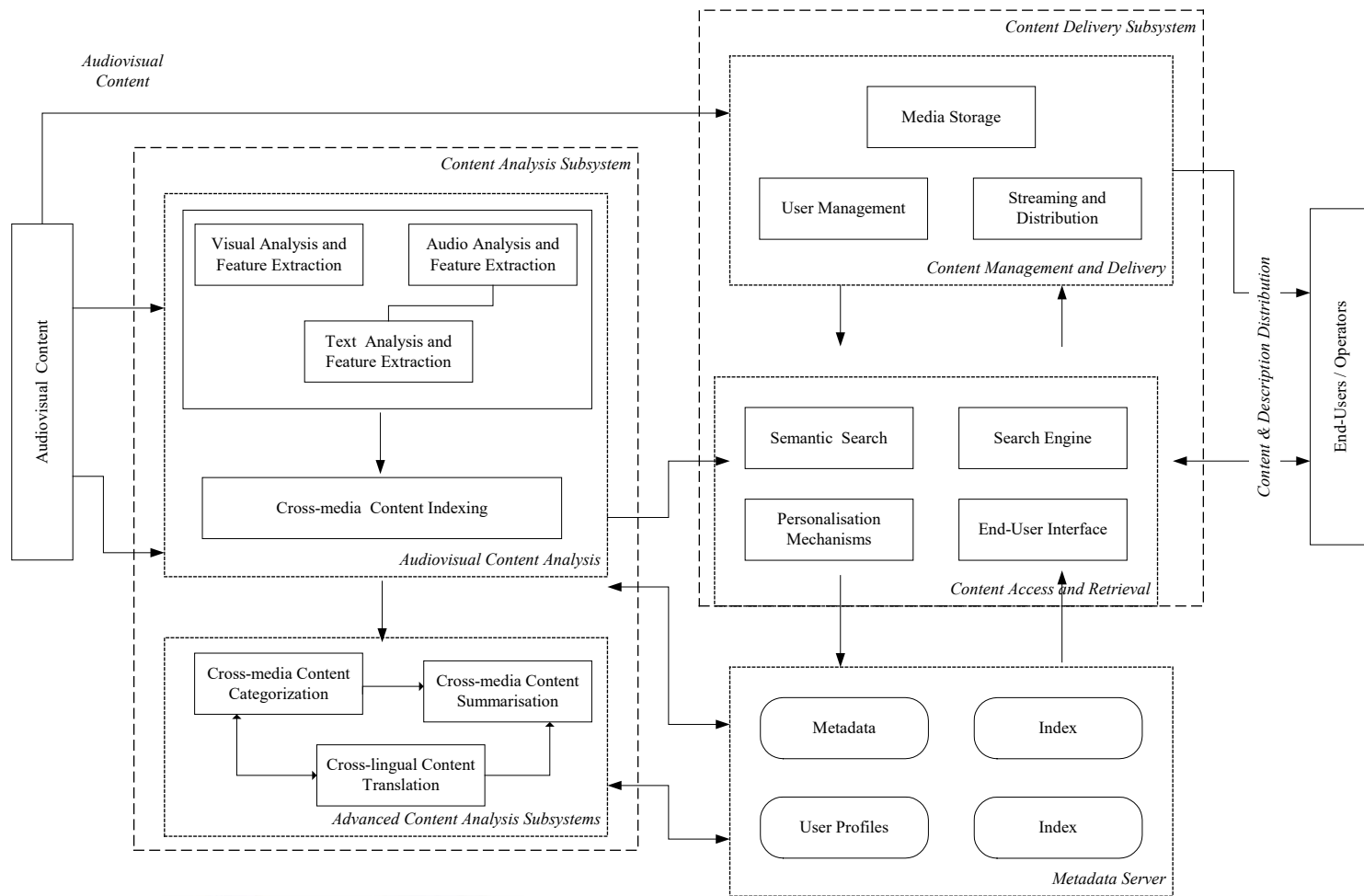
Domaines

- Information
- Documentation automatique
- Catalogage
 - Archives
 - Contenus entrants
- Fournisseurs de contenu traditionnels: diffuseurs, production, publication, publicité
- Nouveaux entrants dans les médias: DVR, VoD, internet à haut débit
- Nouveaux fournisseurs de contenu: financiers, ventes, services, paris
- Fournisseurs de contenu internet: webcasters, portails
- Opérateurs de réseaux: DSL, modem câble, 2G/3G mobile
- Equipementiers: appareils, logiciel, e-commerce, plate-formes d'accès

Standards

- MPEG-7
- MPEG-21
- SVG
- SMIL
- the Corpus Encoding Standard (CES) of the Text Encoding Initiative
- the XML standardization efforts of the W3C such as XLink and XPointer
- the semantic metadata markup languages such as RDF, Topic Maps and TimeML

System architecture



Topic: Colombia drug war

2005-11-17



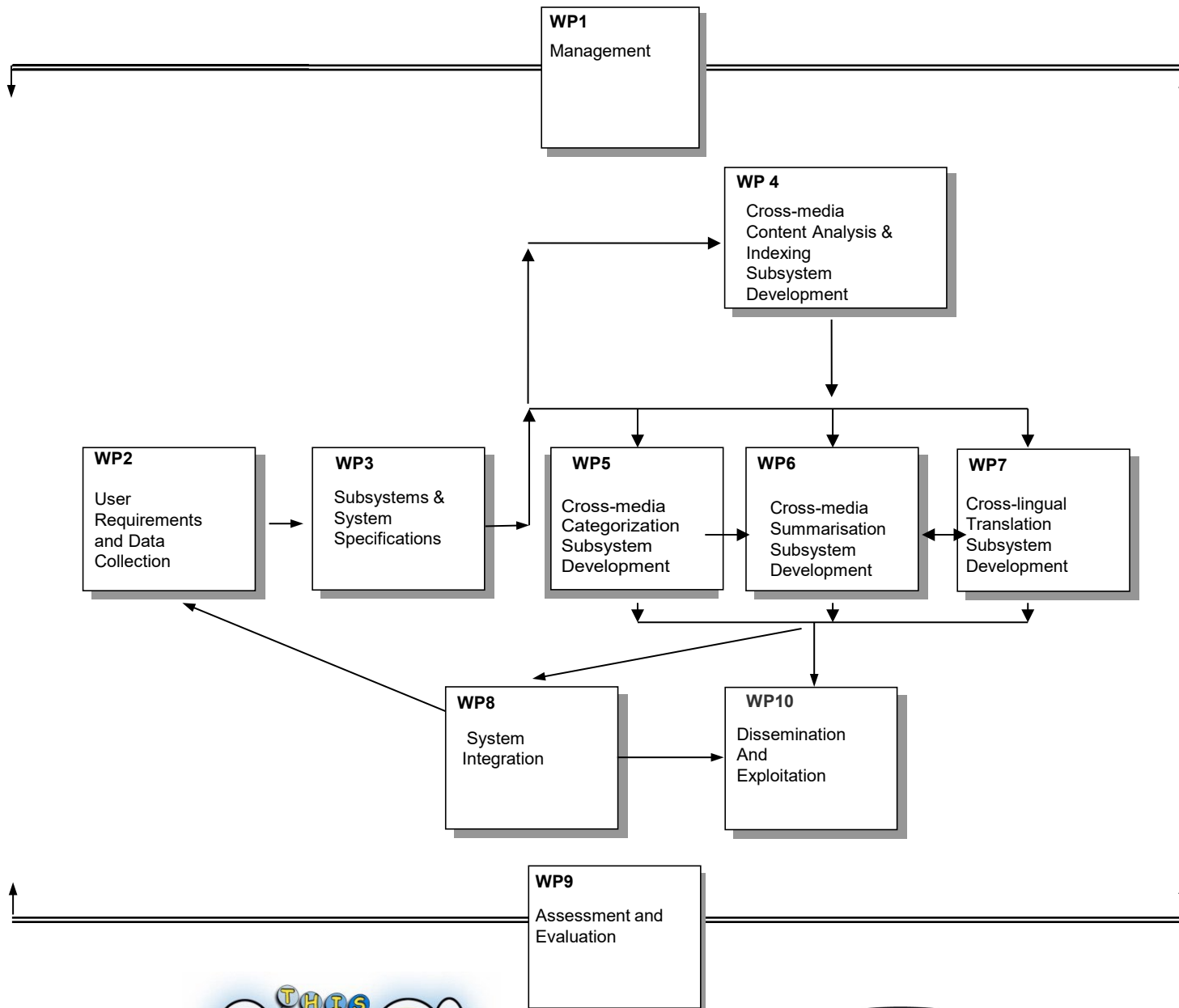
Francis BODSON



15

Logique du projet

- Durée 30 mois
- 10 groupes de tâches
- 33 documents à délivrer



Besoins

- Données audio annotées (Français principalement)
- Données vidéos annotées dans des domaines variés (voyages, politique, ...)
- Testeurs
 - Annotations de sujets
 - Générateurs de questions
 - Evaluateurs

Conclusions

- Projet ambitieux
- Destiné à couvrir des besoins naissants mais de plus en plus présents
- Formalisation de catalogues