

Université de Tunis II
Ecole Nationale des Sciences de l'Informatique

***Hannibal* - Un Moteur de Recherche**
Avec plug-in en Arabe

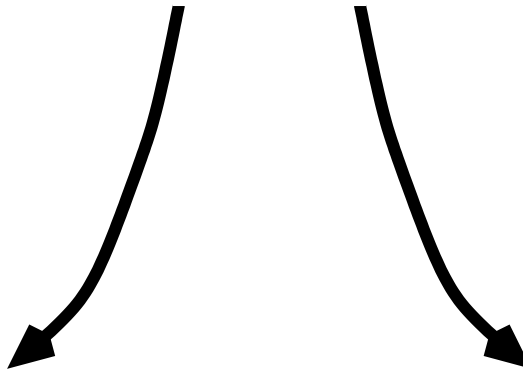
par Imed CHIHI

Encadré par

Mohamed Said OUERGHI, ENSI

Moez SOUABNI, MRS

**Croissance du Web
dans le Monde Arabe**



Accès aux documents
en Arabe

Localisation des
documents désirés

Partie 1: Problématique de la langue Arabe
 Problématique de l'internationalisation en général
 Actions à entreprendre
 Technologie utilisée

Partie 2: Service de recherche
 Fonctionnalités

 Conclusion & perspectives

Catégories des problèmes

- 1.** représentation des caractères
- 2.** inconsistences causées par la multiplicité des langues de saisie
- 3.** satisfaire les besoins sans s'isoler

Multitude des jeux de caractères

Méthodes de saisie peu élaborées

Morphologie complexe

Physiologie complexe

Bidirectionnalité

US-ASCII était *le* jeu de caractères

Nouvelles tendances:

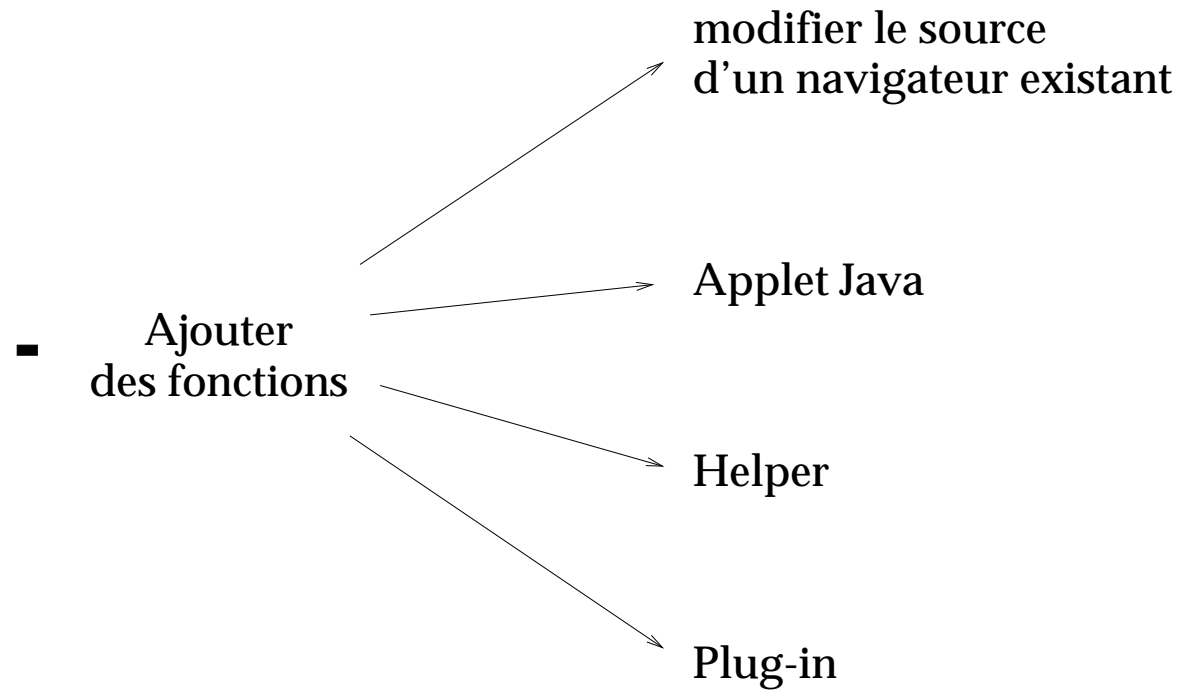
- ISO/IEC 10646, puis Unicode,
- Améliorations de HTML,
- Utilisation de MIME avec HTTP,
- ISO 10646 jeu de caractères privilégié,
- RFC 1556 pour la directionnalité.

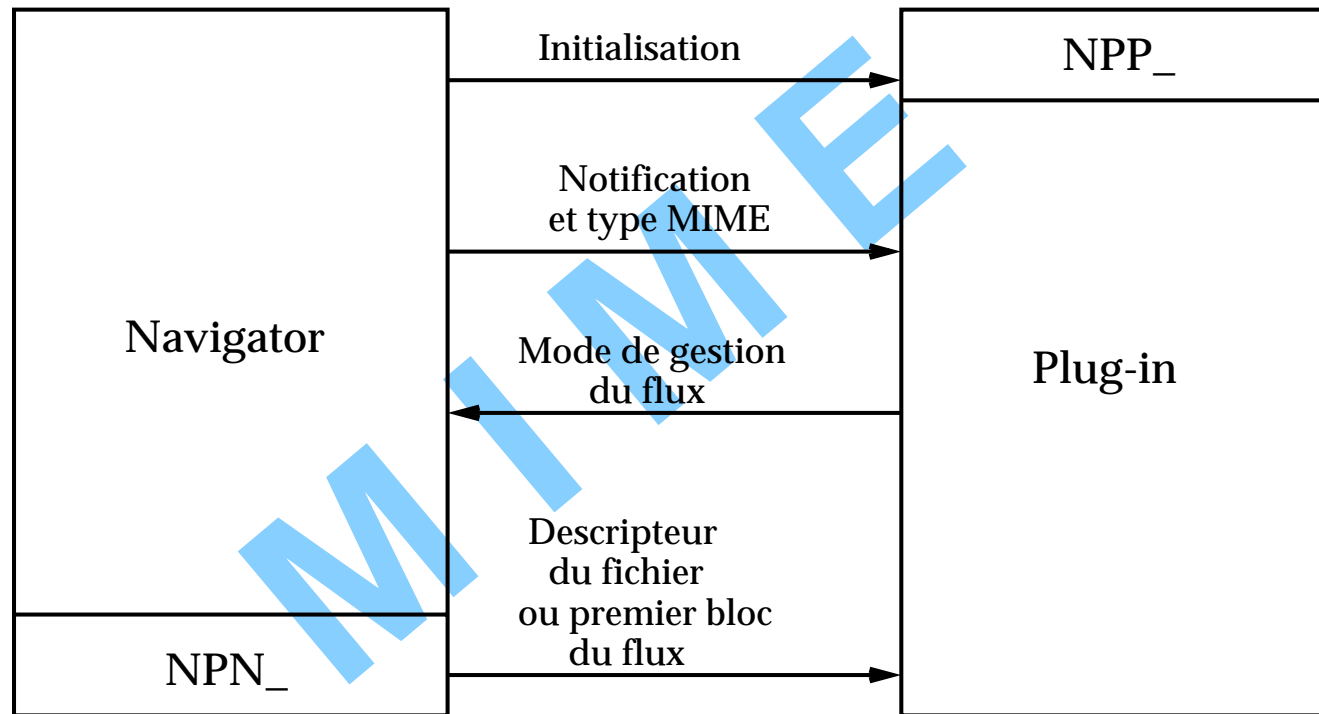
➔ Nécessité d'agents utilisateur plus *intelligents*

Agents utilisateur

	CP 1256	ISIRI 3342	ISO 8895-6	UTF	Unix	Mac	Windows
Sakhr Sindbad			X	X			X
Alis Tango	X	X	X	X			X
Accent Multilingual Mosaic	X		X				X
Netscape Communicator				X	X	X	X
LangBox AraMosaic	X	X	X		X		
PMosaic		X			X		
Internet Explorer				X	X	X	X

- "Compter" sur Unicode





text/html; charset=iso-8859-6

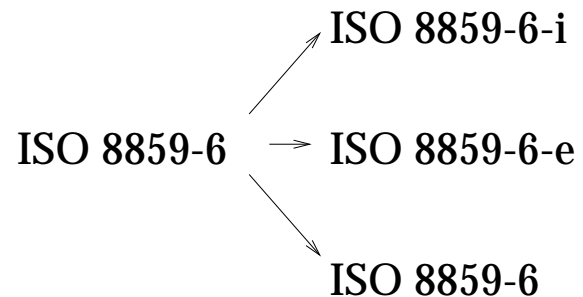
- plug-in trop volumineux,
- n'est pas l'objectif du sujet,
- solution existante

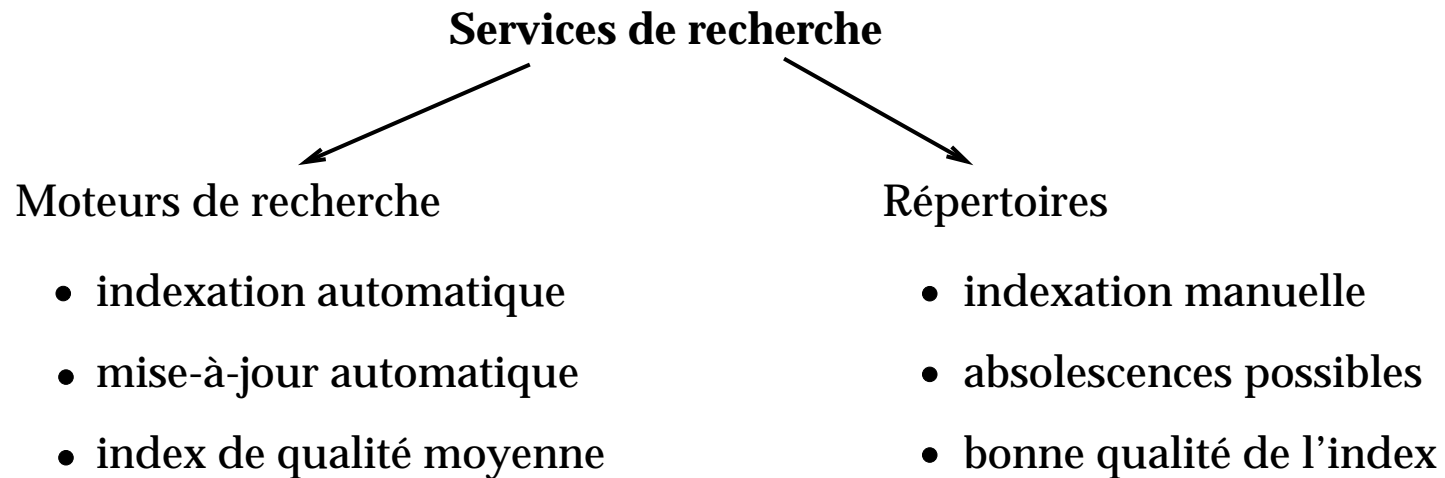
→ text/plain: texte non formaté

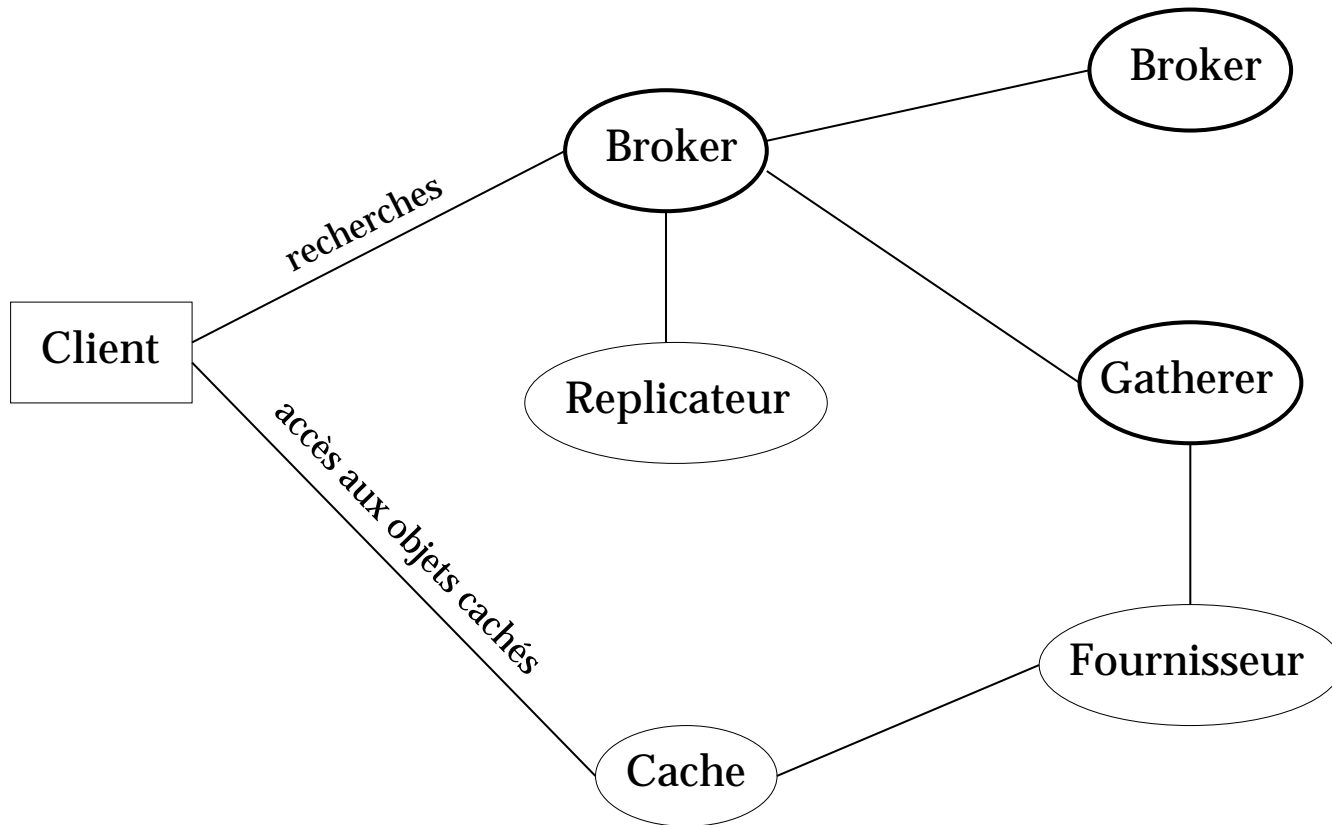
Jeu de caractères:

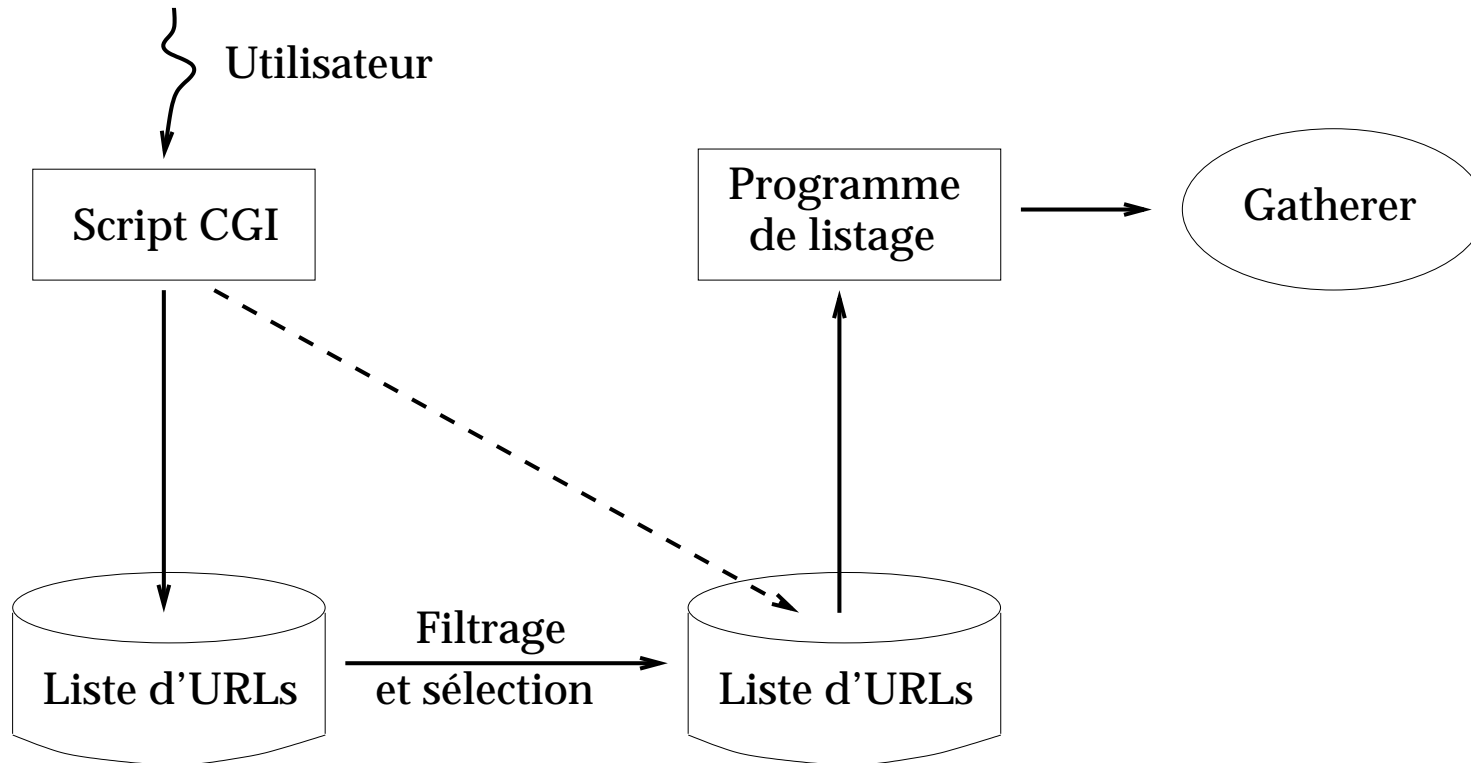
CP 1256 propriétaire

ASMO_449 pas de glyphes Latins
et ISIRI 3342









Respect de la casse,

Accepter les mots clés augmentés,

Supporter des erreurs dans les mots clés,

Affichage des lignes d'apparition des mots clés.

- Ce qui a été fait:
 - Etude des standards et technologies existants,
 - Outil de visualisation de documents Arabes,
 - Service de recherche fonctionnel.

- Solution ultime pour les problèmes de l'Arabe:

Unicode

— Plug-in

- portage vers d'autres plate-formes,
- ajout d'autres encodages (surtout UTF-8),
- promouvoir les systèmes auteurs en Arabe.

— Service de recherche

- distribuer les processus de gathering et de brokering,
- indexer des documents en Arabes.