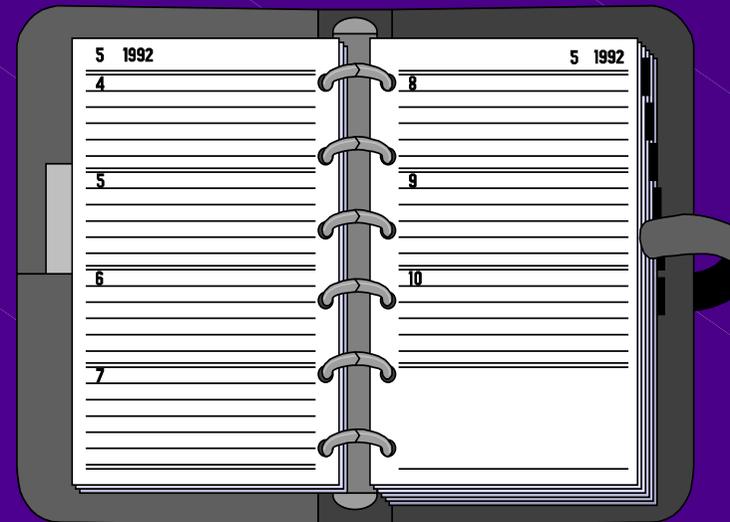


# WAIS

- Wide Area Information Servers



[cours@urec.cnrs.fr](mailto:cours@urec.cnrs.fr)

# WAIS

- 1994 : Jean-Paul Gautier

# Généralités (1)

- **Développé pour permettre la recherche de documents en posant des questions en langage naturel**
  - » Thinking Machine, Apple Computer, Dow-Jones.
- **Deux notions essentielles et totalement nouvelles par rapport aux autres outils :**
  - » La recherche se fait sur le contenu du document, le texte qui le compose et non plus sur les titres et les mots-cles.
    - recherche dite en "texte intégral"
  - » Classement des documents trouvés par ordre de pertinence  
=> la pertinence de la saisie des critères conditionne la pertinence de la recherche.

## Généralités (2)

- **Ensemble de logiciels**
  - » création de bases de données indexées sur des serveurs accessibles par l'Internet.
    - base WAIS, source WAIS
- **Le dialogue client-serveur s'appuie sur une variante du protocole Z39.50 de l'ANSI**

# Les logiciels WAIS (1)

- **Un programme d'indexation : waisindex**
  - » construction d'un index par balayage d'un ou plusieurs fichiers.
  - » l'index contient une ligne de description pour chaque document : Header Line
  - » l'index comporte une entrée par mot
  - » une entrée pointe vers une liste de documents avec chacun un poids
    - poids proportionnel au nombre d'occurrences du mot dans le document.
    - indexation paramétrable.

# Les logiciels WAIS (2)

- **Un programme serveur : waisserver**
  - » répond à une requête par une liste ordonné de "Header Lines"
  - » à une demanade de transfert, il envoie le document
  - » format du document:
    - texte ASCII, Postscript, dvi, images GIF, TIFF, documents MIME
- **Un programme client**
  - » Pose une question à un où plusieurs serveurs
    - l'opérateur implicite est le OU
  - » Unix : waissearch, xwais,..
  - » Mac : WAIStation, PC : oacwais, wwais103

# Base WAIS (1)

- **Un fichier suffixé par .src décrit la base**
  - » nom de l'index, nom du serveur, numéro de port TCP.
  - » la description du contenu de la base.
  - » les coordonnées de l'administrateur.
- **Mise à jour**
  - » stockage des documents dans le même format que les précédents.
  - » régénération du fichier d'index.

## Base WAIS (2)

- **Une base mondiale existe**

- » regroupe l'ensemble des fichiers .src des bases existantes sur l'Internet : environ 700 bases
- » constitue l'annuaire mondiale des bases WAIS : directory of servers
  - machine quake.think.com
- » annuaire des bases françaises : directory-zenon-inria-fr
  - machine zenon.inria.fr

- **Wais peut frontaliser un autre SGBD**

- » nécessite un traducteur
- » ajout du résultat dans de nouveaux documents
- » refaire une indexation WAIS
  - exemple : TEXTO

# Interrogation

- **Première utilisation sur un sujet**
  - » interrogation de l'annuaire mondiale avec des mots clés.
  - » réception d'une liste de bases traitant du sujet.
  - » ensuite on peut lancer des interrogations sur les bases
- **Création d'un environnement avec trois répertoires**
  - wais-sources : fichiers ;src des sources intéressant l'utilisateur
  - wais-questions : texte des questions prêtes à l'emploi
  - wais-documents : documents sauvegardés lors d'interrogations précédentes

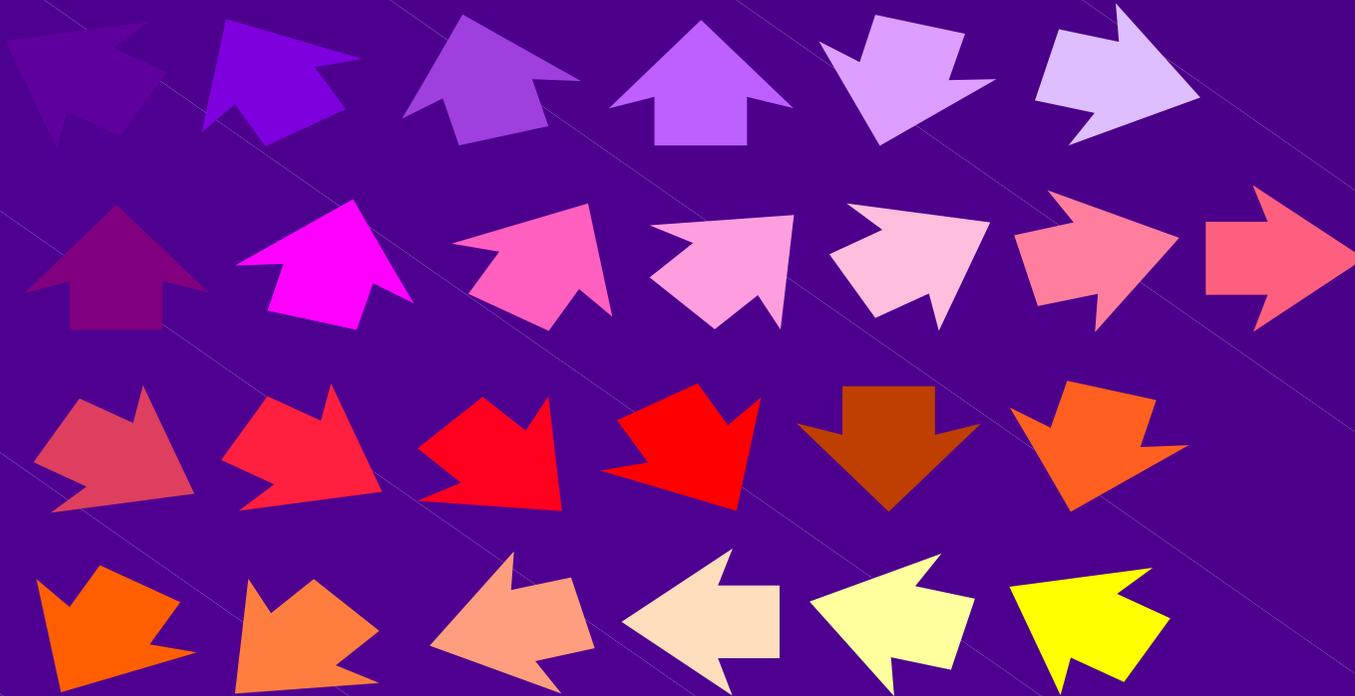
# Limites de Wais (1)

- **Le monde de l'information est représenté par une structure plate : la base**
  - » pas de pointage vers une autre base
- **L'annuaire n'est pas un outil suffisant pour accéder rapidement à ce que l'on cherche.**
- **On atteint les documents par mots-clés**
  - » on ne progresse pas dans une arborescence de fichiers ou dans de l'Hypertext.
- **La syntaxe des requêtes est pauvres.**

## Limites de Wais (2)

- **WAIS n'est pas reconnu comme un gestionnaire de base de données**
  - » absence de thesaurus, classification très simple des documents
- **Wais est utilisé pour indexer les fiches bibliographiques de fonds documentaires, des rapports de recherche, des publications.**

# GOPHER



Jean-Paul GAUTIER

# Généralités

- **Conçu par l'Université du Minnesota**
  - » service d'information de campus
- **Application Client-Serveur avec un protocole de communication spécifique.**
- **Instrument de navigation**
  - » voyage entre tous les serveurs de l'Internet
  - » récupération de documents sur l'Internet
- **Les documents sont présentés sous forme d'une arborescence similaire au système de fichiers Unix.**
  - » promenade dans l'arborescence.
  - » interrogation de bases de données indexées avec des mots clés.
  - » visualisation d'un document.
  - » transfert et stockage d'un document.

# Logiciels

- **serveurs**

- » Unix : démon gopherd
- » Mac, VMS

- **clients**

- » Unix : gopher, xgopher
- » Mac : TurboGopher, GopherApp
- » PC : PcGopher, GopherBook

# Mode opératoire : l'utilisateur

- **Au lancement d'un client :**
  - » connexion automatique à un serveur, le "home server".
  - » envoi par celui-ci d'une page d'accueil.
    - liste de répertoires, de fichiers, d'index.
- **En pointant sur un répertoire, on affiche une nouvelle liste.**
- **En pointant sur un fichier ou un index, selon son type :**
  - » affichage du contenu du document (ASCII, Postscript).
  - » visualisation d'une image (format GIF..)
  - » restitution d'un enregistrement sonore.
  - » transfert et stockage sur un disque local (.Z, .tar. ..)
  - » demande de mots clés (index)
  - » ouverture d'une session telnet
- **Accès à l'information de manière transparente pour l'utilisateur dans tout l'Internet : "Gopher Space"**

# Mode opératoire : le protocole

- **Un serveur écoute les requêtes sur un port TCP connu**
  - » en général le port 70
- **Echange de pointeurs entre le serveur et le client**
  - » a chaque requête du client, le serveur retourne une liste d'éléments qui contiennent des pointeurs.
- **Les champs d'un élément**
  - » **TYPE** : qualifie le document présenté
    - 0 texte, 1 répertoire, 3 message d'erreur, 4 fichier BinHex pour Mac
    - 5 binaire DOS, 6 fichier "uuencodé", 7 index , 8 session telnet
    - 9 fichier binaire, s fichier audio, l image GIF
  - » **NOM** : chaîne de caractères à afficher identifiant la ressource.
  - » **SELECTEUR** : localisation du document sur la machine.
  - » **MACHINE** : identité de la machine qui héberge le serveur.
  - » **PORT** : numéro de port TCP du serveur référencé.

# Les passerelles

- **Gopher permet d'interroger des services qui n'utilisent pas son protocole au moyen de passerelles qui présentent les résultats sous forme de menus Gopher.**
  - » ftp anonymous
  - » WAIS
- **Une toile d'araignée relie tous les serveurs gopher.**
  - » dans chaque pays un site tient à jour la liste des serveurs du pays : en France [gopher.jussieu.fr](http://gopher.jussieu.fr)
- **VERONICA : service qui gère une base indexée du contenu des serveurs gopher mondiaux.**
  - » interrogation par mots clés
  - » récupération d'un répertoire dont les éléments sont des pointeurs vers des serveurs gopher qui contiennent des documents en rapport avec la requête.

# Avantages - Limites

- **Interface conviviale, transparence pour l'utilisateur.**
- **Vision hiérarchique de l'information.**
- **Facilités d'installation et de gestion.**
  - » identique au travail de gestion d'un serveur ftp anonyme.
  - » il est facile de structurer un serveur gopher pour une population dont on connaît les besoins.
- **Accès à des documents distants grâce aux pointeurs.**
- **Nécessité d'un logiciel client sur chaque poste de travail.**
- **Les passerelles sont dans le serveur et non dans le client.**
- **N'offre pas les possibilités de l'hypertext et un format de document standard.**