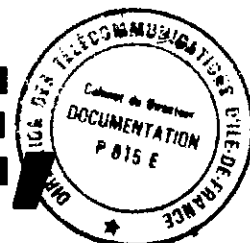


# MINITEL GUIDE L'ANNUAIRE INTELLIGENT



*Revue du Minitel*

M: 8, Novembre -

Décembre 1986,

pp. 35-40

MANO SIRI

*Comment se retrouver dans la jungle des services ? Les annuaires actuels, qu'ils soient électroniques ou papier s'avèrent peu efficaces. Un nouveau service en cours de mise en place, « Minitel guide », devrait combler ce manque. Il bénéficiera, dès 1987, des derniers développements de l'intelligence artificielle : on pourra l'interroger en français courant, sans contraintes.*

**Q**UI ne connaît l'annuaire électronique ? Aujourd'hui encore, c'est le service Télétel le plus souvent consulté : au hit-parade des services, il caracole en tête, loin devant les services leaders du 36.15, et a marqué de son sceau l'avènement de la télématique. Tout nouveau possesseur d'un Minitel fait connaissance avec le vidéotex à travers la logique de l'utilisation de l'annuaire. Il y découvre ce qui fonde le pari de la télématique : une simplicité de dialogue, de la conversation homme machine qui ne nécessite aucun apprentissage, aucune compétence informatique particulière pour formuler et faire aboutir sa demande.

Dès le premier écran, l'utilisateur n'a plus qu'à remplir les zones « nom » et/ou « rubrique », « localité » et, s'il le souhaite, préciser sa recherche en spécifiant le « département », l'« adresse », le « prénom ». Le système recherche automatiquement la réponse correspondant à sa demande ; s'il ne trouve pas, il effectue une recherche phonétique sur les noms propres ou les noms communs qui permet de retrouver la réponse par approximation, mais peut aussi engendrer beaucoup de « bruit » dans les réponses. En cas d'absence de réponse, il propose à l'utilisateur un « élargissement » de sa recherche : sur une autre adresse, un autre prénom ou des localités voisines de celle demandée...

Mais cette sophistication dans l'analyse de la demande — qui fait déjà figure de modèle pour les autres services télématiques — montre pourtant ses limites dès qu'il s'agit d'interroger l'annuaire par professions...

Par exemple : si l'utilisateur tape « constructeur de planches à voile », le système actuel lui propose 16 rubriques :

1. Chaudronnerie (constructeurs).
  2. Machines (constructeurs).
  3. Fours (constructeurs).
  4. Aéronautique (constructeurs pour).
  5. Grues (constructeurs).
  6. Maisons individuel les (constructeurs).
  7. Promoteurs (constructeurs).
  8. Puits (constructeurs).
  9. Radio-électricité (constructeurs d'appareils).
  10. Bateaux de plaisance et planches à voile (construction et agencement).
  11. Bateaux de plaisance et planches à voile (location et vente).
  12. Bateaux de plaisance (achat et vente).
  13. Planches à voile.
  14. Location de bateaux, canoës kayaks et planches à voile.
  15. Bateaux de plaisance et planches à voile (matériels et fournitures pour).
  16. Location de bateaux de plaisance, canots et planches à voile.
- Seule la 10<sup>e</sup> rubrique correspond à la

demande... Il apparaît ainsi que le système ne sait ni distinguer d'un point de vue sémantique le terme « constructeur » du terme « planche à voile », ni combiner ces deux notions pour en faire une unité de sens. C'est cette analyse sémantique et syntaxique que se propose et parvient à faire l'automate de langage naturel développé par ERLI.

Dans l'exemple précédent, pour accéder directement à sa réponse, l'utilisateur aurait dû taper « bateaux de plaisance et planches à voile (construction et agencement) » : cela suppose qu'il connaisse le libellé exact

La logique de recherche documentaire en télématique grand public et professionnelle est passée par plusieurs phases d'optimisation du dialogue et de l'interrogation homme-machine :

1. La logique arborescente : l'utilisateur affine sa demande à travers une succession d'écrans en répondant aux questions préétablies par le système.

Dans le cas d'une demande précise, la recherche risque d'être très longue avant d'aboutir à une réponse qui ne satisfera peut-être pas l'utilisateur. Et la démarche adoptée représente la logique préétablie par le système : ce

d'écrans correspondants à un parcours préétabli.

4. Certains « multicritères » sont parfois doublés de thesaurus ou dictionnaire de mots appartenant au vocabulaire de l'utilisateur. Le système peut alors reconnaître la demande de l'usager : un même mot peut alors se décliner au pluriel et au singulier ; il peut aussi avoir des synonymes ou des mots dont le sens lui est associé ; il peut enfin appartenir à un « genre » ou à un ensemble de mots plus vaste. Par exemple la « SYMPHONIE PASTORALE » aura pour synonyme « PASTORALE » ou « LA 5<sup>e</sup> », pour mot associé « BEE-THOVEN », et elle appartiendra au genre « SYMPHONIE ».

Il y a là une tentative d'analyse de la demande de l'utilisateur : le système essaye de gérer celle-ci au plus près de l'univers sémantique de l'usager. Mais un mot, ici, est encore une simple chaîne de caractères : il n'est pas vraiment intégré dans une chaîne de sens, il n'a pas de place dans une phrase ordonnée par une syntaxe particulière.

La logique d'un tel système — qui ne fait aucune analyse linguistique et ne dispose pas de véritables réseaux sémantiques — nécessite d'enrichir à l'infini le dictionnaire des termes ou expressions permettant l'accès à l'information et risque de pénaliser l'utilisateur en temps de réponse.

## QUI EST ERLI ?

*ERLI, société d'étude et recherche en linguistique et informatique, existe depuis 1977.*

### *Philosophie générale :*

*ERLI a pour vocation l'étude et la réalisation de systèmes mettant en oeuvre les techniques de la linguistique automatique et de l'intelligence artificielle appliquées aux systèmes d'information. Cette activité est plus particulièrement orientée vers la conception de systèmes d'information et bases de données, et la réalisation de logiciels facilitant le dialogue homme/machine : analyse du langage naturel, télématique grand public.*

### *Les produits et les services offerts par ERLI :*

• le logiciel SAPHIR : un analyseur du langage naturel adapté à l'interrogation de bases de données factuelles (gestion du personnel, commercial,...).

Ce logiciel est disponible sur matériel IBM sous MVS et VM/CMS.

• Le logiciel ALEXIS : il est destiné à la gestion de dictionnaires complexes, permettant de constituer des bases de données de mots ou de termes.

Ce logiciel est disponible sur les matériels IBM sous MVS et VM/CMS.

• Etudes et réalisations.

des rubriques professionnelles contenues dans l'annuaire, qu'il intègre la nomenclature préétablie par les concepteurs de l'annuaire. Dans un système en langage naturel, l'utilisateur peut parvenir, immédiatement ou presque, à la rubrique qui l'intéresse. La Direction générale des télécommunications (DGT) a souhaité améliorer l'interrogation des rubriques professionnelles dites « pages jaunes » de l'annuaire. Il s'agit d'offrir un service répondant le mieux possible aux besoins des usagers : à partir d'une demande exprimée dans le langage et le vocabulaire de l'utilisateur, le système doit pouvoir trouver et proposer la ou les rubriques professionnelles correspondantes. L'utilisateur n'est pas censé connaître le libellé des rubriques — on en compte 3 500 — qui composent la nomenclature de l'annuaire.

La DGT a donc confié à ERLI l'étude et la réalisation d'une maquette, puis d'un prototype d'interrogation en langage naturel qui doit être prochainement intégré dans l'architecture générale de l'annuaire.

n'est pas celle, « naturelle », de l'utilisateur.

2. La recherche par mots clés : plus souple, elle laisse plus de latitude à l'utilisateur... Mais si le mot entré est écrit au pluriel, alors qu'il est stocké dans le système au singulier, il arrive assez souvent qu'aucune réponse ne puisse être apportée à l'utilisateur. Celui-ci reçoit alors un message du type « mot clé inconnu » ou « il n'y a pas de réponse à votre demande ».

3. Le multicritère : l'utilisateur remplit une ou plusieurs zones en indiquant les mots ou les expressions qui spécifient sa demande. Le système fait alors l'intersection entre les différents critères et propose à l'utilisateur le résultat de sa recherche.

Exemple : 1<sup>er</sup> critère, SYMPHONIE PASTORALE ; 2<sup>e</sup> critère, BARENBOIM ; 3<sup>e</sup> critère, 15-12-86. Si le système ne trouve rien, il redonne la main à l'utilisateur. Après avoir reçu une première question ouverte, certains systèmes n'offrent plus d'autres possibilités que celle de se laisser guider à travers une succession

## MINITEL FICTION

*Dis-moi ce que tu veux, je te dirais à qui t'adresser... Ainsi pourrait-on résumer les quelques essais de recherche effectués sur l'annuaire à l'aide de l'automate développé par ERLI :*

« j'ai faim » donne :

- restaurant
- ALIMENTATION générale
- coopérative d'alimentation.
- « où dîner ce soir » propose :
- restaurant
- cabaret
- music-hall.

« transporter des pêches par avion » donne : transport aérien ;

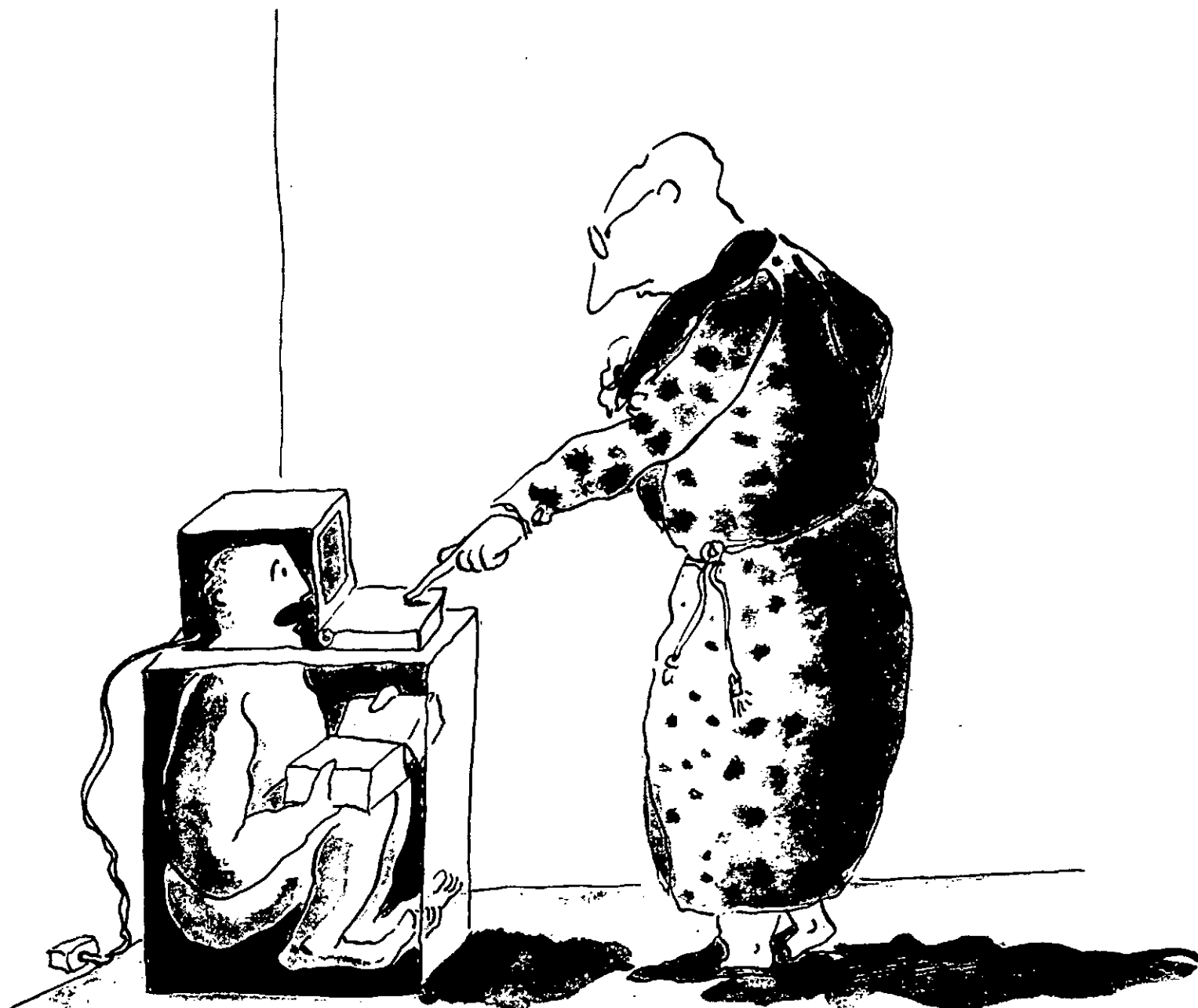
« aller à la pêche » donne : pêche et chasse (société).

« acheter des articles de pêche » aboutit à :

- pêche et chasse (articles de)
- pêche et pisciculture (matériel et fournitures)

« confesser mes péchés »,

le système me propose : Eglise catholique.



Dans le nouveau système, les termes sont gérés dans un dictionnaire incluant un réseau de relations précisant leur environnement sémantique: cf. graphique n° 1. Le système corrige les fautes de frappe et d'orthographe, tient compte des mots composés (une *armoire à pharmacie* ne peut être considérée comme l'équivalent de *armoire* et *pharmacie*), de l'ordre des mots (*chevaux de course* et *courses de chevaux*, par exemple), de leur catégorie (certains mots peuvent être soit des noms, soit des adjectifs, selon le contexte) et surtout de la sémantique (*louer une voiture* doit pouvoir renvoyer à *location d'automobiles* et de *camions*). Cela suppose qu'un mot ne soit plus une simple chaîne de caractères :

il apparaît dans un contexte, dans une phrase qui n'est pas une simple suite non ordonnée de mots. Le système d'ERLI est donc constitué d'un dictionnaire et de règles syntaxiques permettant de reconnaître la structure d'une phrase, et d'un automate qui, schématiquement, doit gérer ces deux outils linguistiques.

Le processus général d'analyse est le suivant: il peut être décomposé en plusieurs phases d'analyse de la requête de l'utilisateur.

1<sup>re</sup> phase: l'analyse lexicale et morphologique. Les mots simples ou composés sont tout d'abord reconnus et il leur est affecté un type morphologique: l'automate doit ainsi distinguer les noms, les adjectifs, les verbes, les

propositions... Dans certains cas, il y a analyse d'ambiguïtés morphologiques: selon le contexte, «informatique» pourra être nom ou adjectif, par exemple. Cette première phase permet de reconnaître en tant que tel les mots composés: producteur de pommes de terre; stage de ski de fond; réparation de pots d'échappement.

Cela est évidemment primordial lorsque le sens du mot composé n'est pas fonction directe de chacun de ses composants: une pomme de terre n'est pas une sorte de pomme par exemple... L'analyse morphologique permet également l'élimination des mots inconnus, le redressement orthographique si besoin est, ainsi que la mise à l'écart de tous les mots qui n'apportent pas

# EXPLOITEZ

votre futur

## CENTRE SERVEUR

# RÉALISEZ

sur

## NOS SYSTÈMES UNIX

### VOS APPLICATIONS

A votre disposition :

- Générateur d'écrans videotex
- Gestionnaire d'arborescence
- Compta usager
- Messageries
- Fichiers multicritères
- + Nos locaux
- + Nos techniciens

**TÉL. (1) 45.39.10.90**

Protection des projets  
par  
Huissier de Justice

#### DIALOGUE

d'information dans l'analyse de la requête. Par exemple « je voudrais », « je cherche », « où trouver », « quel est »...

2<sup>e</sup> phase : l'analyse syntaxique. Elle utilise la liste d'expressions reconnues comme étant pertinentes grâce à l'analyse précédente, pour reconnaître la structure de la question de l'utilisateur et en déterminer les points forts. L'analyse de la requête de l'usager permet ainsi de constituer des couples « prédicat-thème » dont l'existence est vérifiée dans le dictionnaire. Le prédicat correspond à ce qui est affirmé à propos du thème. Ainsi, « automobile, bateau » sont des objets ou thèmes sur lesquels l'utilisateur peut demander d'appliquer des prédicats

tels que « acheter, louer, réparer... » Un même prédicat peut être exprimé de différentes manières : ainsi les requêtes : « achats de voitures » ; « concessionnaires de voitures » ; « acheter voiture », seront équivalentes, les prédicats faisant partie de la même famille sémantique.

De même, sont reconnues comme équivalentes les différentes formes suivantes : « ma voiture est en panne ; faire réparer ma voiture ; réparation de voiture ; voiture à réparer ; réparateur de voiture ; panne de voiture ».

Le mot « voiture » est relié à la rubrique « Garage » de l'annuaire par la relation prédictive représentant l'idée de « réparation ». L'analyse syntaxique permet en outre d'identifier les rapports entre les mots et de décider quelles associations privilégier. Si la requête de l'utilisateur est : « Je cherche un vieux meuble en chêne », l'automate cherche d'abord à résoudre « vieux meuble en chêne » et s'il ne trouve aucune rubrique correspondante, il cherchera alors « vieux meuble » et « meuble en chêne », mais ni « vieux chêne » ni « vieux » ne peuvent constituer des expressions pertinentes. Vieux est, dans ce contexte, un adjectif et ne doit pas être analysé seul.

3<sup>e</sup> phase : l'analyse sémantique. Elle a pour but l'affichage des résultats, c'est-à-dire des rubriques professionnelles correspondant à la question de l'utilisateur.

Tout thème est considéré comme un index possible des rubriques profes-

sionnelles et l'annuaire. Il s'agit donc de faire correspondre la question initiale de l'utilisateur avec un index existant dans la base de connaissances. L'automate d'ERLI opère des « substitutions lexicales » : il répondra aussi bien à « Achat d'auto » qu'à « Achat d'automobiles » ; il procède également à des transformations syntaxiques. Le réseau sémantique et syntaxique qui

organise les termes de la question entre eux permet de retrouver les rubriques correspondantes dans les pages jaunes de l'annuaire. La question « je veux faire réparer un fauteuil ancien » peut bien ne renvoyer directement à aucune rubrique ; l'automate transforme la question initiale en liant « fauteuil » à « meuble », ce qui

**“ Tapez au clavier : « ma batterie est en panne... » et la liste des garages les plus proches apparaît sur votre écran ! ”**

donne « réparer un meuble ancien » et permet d'identifier la rubrique « Ebénistes ». Le choix des relations sémantiques entre les mots est guidé par le contexte :

- Ma batterie est en panne. Donne :
- Garages automobiles.
- Electricité pour automobile.

Mais :

- Batterie de cuisine aboutit à :
- Quincaillerie ;
- Articles de cuisine.
- Je veux apprendre à jouer de la batterie est relié à :
- Musique et chants (leçons) ;
- Conservatoires.

Ce système permet ainsi de répondre à des demandes aussi variées dans leur formulation que dans leur objet telles que : Voiture réparation ; Je voudrais réparer ma voiture ; Je cherche un stage de ski ; Louer un studio ; Louer un studio photo ; Transporter un bateau par camion ; Transporter un camion par bateau ; Mon enfant est malade. L'ouverture réelle de ce service, qui peut paraître magique, voire ludique, est prévue début 1987 et devrait permettre d'en mesurer plus précisément l'intérêt pour l'utilisateur : une évolution possible du système est d'ailleurs envisageable selon ERLI, fonction des demandes de l'utilisateur, de son appropriation de ce système qui peut avoir d'autres applications télématiques telles que l'interrogation des petites annonces, la recherche de programmes de cinéma, d'informations touristiques... Autrement dit : A vous de jouer !