IDIAP

Rapport technique

Une interface d'indexation documentaire I d'i, version 1.4

Jean-Luc Cochard

Février 1993

INSTITUT DALLE MOLLE D'INTELLIGENCE ARTIFICIELLE PERCEPTIVE
CASE POSTALE 609 - 1920 MARTIGNY - VALAIS - SUISSE
TELEPHONE : ++41 26 22.76.64 - FAX : ++41 26 22.78.18
E-MAIL : IDIAP@IDIAP.CH

Numéro : 93-01

Un interface d'indexation documentaire
I d'i, version 1.4

Mode d'emploi

Jean-Luc Cochard

Institut Dalle Molle d'Intelligence Artificielle Perceptive (IDIAP)
Case postale 609, CH-1920 Martigny, Suisse
Adresse électronique : `cochard@idiap.ch`

Résumé

Ce document présente un interface d'utilisation d'un système prototype d'indexation documentaire construit dans le cadre du projet *Specification and Prototyping of a System for the Intelligent Management of Information*[1]. Cet interface permet de commander l'indexation automatique de documents et de contrôler le résultat du traitement effectué.

D'autre part, cet outil illustre les possibilités d'utilisation d'une boîte à outils graphique, `Tk`, et son intégration possible avec Prolog via un utilitaire appelé `expect`.

Cet interface d'indexation fait partie d'une triade d'outils qui comprend un interface d'administration [Coc93a] et un interface de recherche documentaire [Coc93b] qui font l'objet de rapports distincts.

Mots-clé : interface homme-machine, interface graphique, indexation et recherche documentaire, traitement de la langue naturelle.

[1] Projet N° 4023-26996, PNR 23, FNSRS

Table des Matières

1 Introduction

L'objectif de ce document est de fournir un mode d'emploi aussi complet que possible des commandes disponibles dans cet outil d'indexation automatique de documents : **I d'i 1.4**.

Cet outil permet de faire une indexation structurelle et linguistique de lettres administratives conformément à l'un des objectifs du projet de recherche *Specification and Prototyping of a System for the Intelligent Management of Information* [CHK93] qui a précédé la mise en place de cet outil.

L'indexation structurelle consiste à décomposer le texte d'une lettre en identifiant certaines composantes sensibles comme la *date de rédaction*, le *numéro de référence* et l'*adresse du destinataire*. Ces trois données[2] sont normalisées et constituent l'*index structurel* du document.

L'indexation linguistique est réalisée par un traitement linguistique particulier sur la zone du *sujet* de la lettre. Ce fut le point central du projet de recherche mentionné ci-dessus. Nous vous référons donc au rapport final du projet de recherche pour une description de ce traitement linguistique.

La suite de ce document décrit les différentes opérations implantées dans cet interface d'indexation. Dans la section 2, nous présentons la structure générale de l'interface ainsi qu'un scénario simple d'utilisation de ses possibilités. La section 3 présente l'environnement de sélection de documents. La section 4 explique le fonctionnement de la commande d'indexation et le rôle de la commande de consultation des résultats. La section 5 explique la manière de quitter le programme. La section 6 décrit le format de la commande Unix avec ses paramètres d'exécution, ainsi que les fichiers de description de la langue de l'interface. Et finalement, en conclusion, dans la section 7, nous présentons une liste non exhaustive d'améliorations possibles du produit.

2 Le tableau de bord

L'interface d'indexation documentaire, **I d'i 1.4**, se présente sous la forme d'un “tableau de bord” (cf. Figure 1) constitué de trois zones :

la file de traitement – cette file est symbolisée par une table qui a le titre “**Documents à indexer :**” et qui est assortie d'un *ascenseur*;

les messages d'exécution – tous les messages à l'intention de l'utilisateur sont centralisés dans cette zone;

les commandes – elles sont au nombre de quatre et seront décrites en détail dans les sections suivantes.

L'utilisation de **I d'i 1.4** est relativement simple. Une session standard se déroule de la manière suivante :

1. le lancement de l'interface à l'aide d'une commande Unix (cf. section 6.1);
2. la constitution d'une file de traitement en sélectionnant des documents du système de fichiers (commande “**Sélectionner**...” décrite dans la section 3);
3. le lancement d'une commande d'indexation qui traite séquentiellement tous les documents présents dans la file de traitement (commande “**Indexer**” décrite dans la section 4);
4. la consultation des résultats qui permet de connaître la qualité de traitement atteint par le système en affichant une trace d'exécution (commande “**Résultats**...” décrite dans la section 4);
5. la fin d'une session qui fait suite à une éventuelle itération des étapes précédentes (commande “**Quitter**...” décrite dans la section 5).

[2] La limitation à ces trois éléments d'une lettre est purement arbitraire. Si l'expérience nous montrait qu'un autre élément du texte permette d'améliorer la qualité de la recherche documentaire, il serait tout à fait possible d'étendre la liste des zones sensibles afin d'y inclure ce nouvel élément d'information.

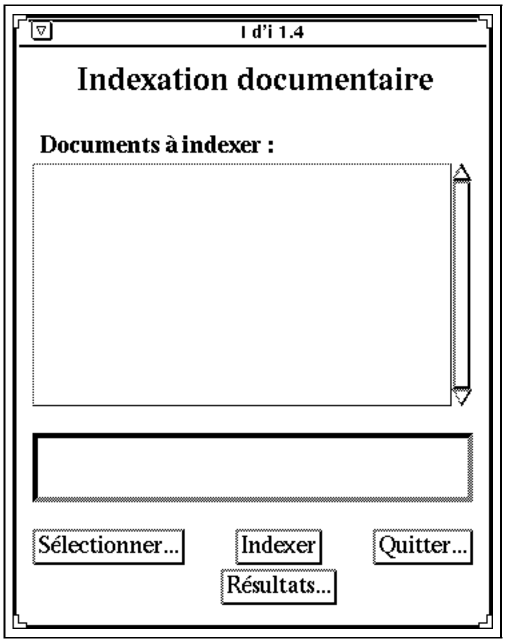


Figure 1 Le tableau de bord de l'interface **I d'i 1.4**.

3 La sélection de documents

La sélection de documents est une opération lancée par la commande "**Sélectionner**...". Elle fait apparaître une fenêtre temporaire (cf. Figure 2) qui permet de se déplacer dans le système de fichiers et de choisir un document.

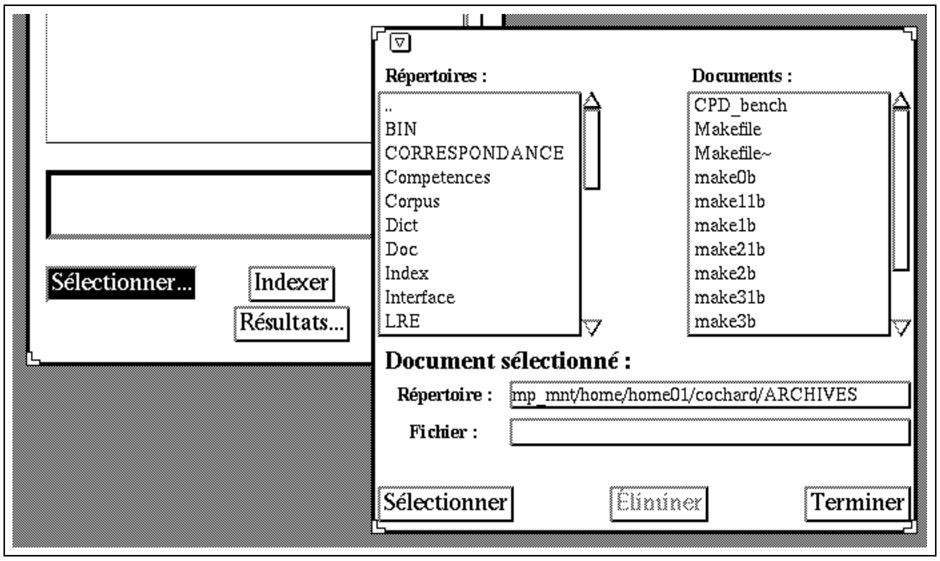


Figure 2 Lors de l'exécution de la commande "**Sélectionner**...", une nouvelle fenêtre temporaire de sélection de documents est affichée à l'écran.

3.1 Le choix du répertoire

Le déplacement dans le système de fichiers permet le positionnement dans la structure arborescente des répertoires. Afin de faciliter la sélection d'un répertoire, deux solutions sont proposées : l'accès contrôlé et l'accès libre.

L'accès contrôlé à un répertoire se fait en utilisant la liste des répertoires sous le titre "**Répertoires :**". Le premier élément de la liste est systématiquement "**..**", qui, selon la convention Unix, dénote le *père* du répertoire courant, tous les autres éléments étant les fils du répertoire courant. Le choix d'un répertoire se fait par sélection de l'élément avec la souris[3]. La sélection d'un nouveau répertoire courant donne automatiquement lieu à une mise à jour de tous les éléments de la fenêtre (cf. Figure 3). L'ascenseur, sur la droite de la liste, permet de se déplacer dans la liste en utilisant indistinctement un des trois boutons de la souris.

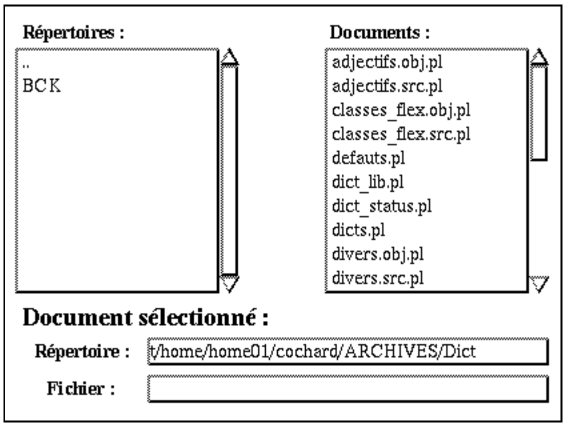


Figure 3 Après avoir sélectionné le sous-répertoire "**Dict**", les différentes composantes de la fenêtre sont mises à jour.

L'accès libre à un répertoire se fait en éditant le contenu de la zone appelée "**Répertoire :**". Les commandes suivantes d'édition de ligne sont disponibles :

Positionnement du curseur d'édition – le positionnement se fait par sélection à l'aide de la souris[4];

Insertion de caractères – l'insertion se fait à la position courante d'édition si le curseur de la souris est positionné dans la zone en question;

Effacement d'un caractère – les touches `Delete` et `Back Space` du clavier effacent le caractère à gauche du curseur d'édition;

Marquage d'une chaîne de caractères – le marquage qui consiste à repérer une chaîne de caractères (affichée en inverse vidéo), est effectué en gardant le bouton de sélection de la souris enfoncée durant un déplacement vers la gauche ou vers la droite;

Effacement d'une marque – la combinaison de touches `Control-D` permet d'effacer une marque;

Effacement du texte – la combinaison de touches `Control-N` permet d'effacer tout le texte de la zone d'édition;

Défilement du texte – lorsque le texte dépasse la taille de la zone d'édition, il est possible de déplacer la portion visible en gardant le bouton du milieu de la souris enfoncé durant un déplacement vers la gauche ou vers la droite.

Lorsque le contenu a été édité, la touche `Return` permet de sélectionner ce répertoire ce qui met à jour l'affichage de la fenêtre.

[3] La "sélection avec la souris" est effectuée en appuyant sur le bouton gauche de la souris lorsque le curseur est sur l'élément en question.

[4] À cause d'un bug de la boîte à outils graphiques, la position courante du curseur d'édition n'est pas toujours affichée.

3.2 Le choix du document

Lorsque le répertoire souhaité est sélectionné, il est possible de faire la sélection d'un document selon le même principe : par un accès contrôlé, via la liste des documents, ou par un accès libre, via la zone appelée "**Document :**". Les commandes d'édition décrites ci-dessus s'appliquent aussi à cette zone.

3.3 La constitution de la file de traitement

L'étape suivante est la transmission des coordonnées du document à la file de traitement. C'est le rôle du bouton "**Sélectionner**" qui ajoute une ligne dans la file de traitement avec le chemin d'accès complet au document (cf. Figures 4.1, 4.2).

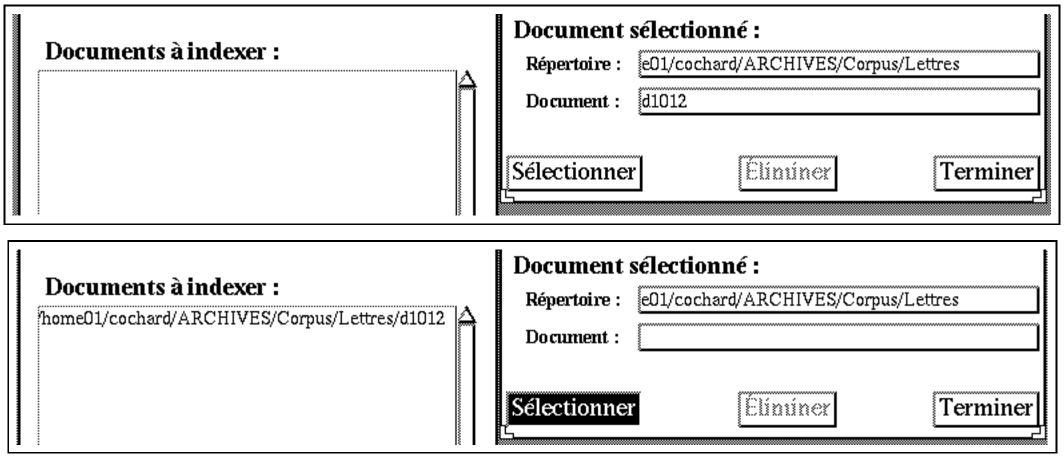
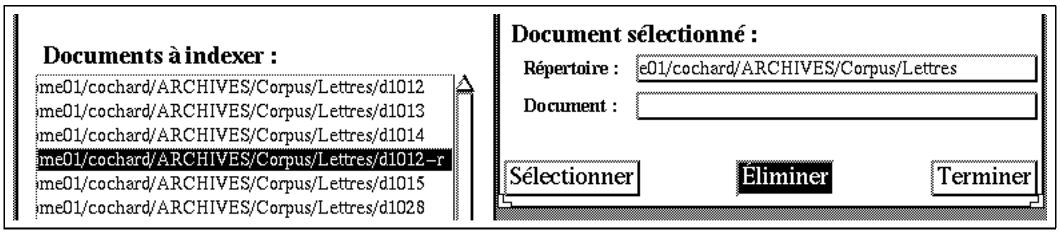


Figure 4 Lorsqu'un document est clairement identifié par son répertoire et son nom, le bouton "**Sélectionner**" effectue le transfert d'information vers la file de traitement. De plus la zone "**Document :**" est effacée afin de permettre la sélection d'un nouveau document.

Pour des raisons ergonomiques, la touche `Return`, tapée n'importe où dans la fenêtre joue le même rôle que le bouton "**Sélectionner**". Comme la technique de sélection de documents adoptée ici ne permet pas de sélectionner une liste de documents par une opération unique, le va-et-vient avec la souris entre la liste des documents et le bouton devient extrêmement désagréable. L'alternative fournie par la touche `Return` permet de garder la souris sur la liste des documents tout en effectuant une transmission dans la file de traitement.

3.4 L'édition de la file de traitement

Comme il n'est pas exclu qu'un mauvais document soit inséré dans la file de traitement, il est possible de l'éliminer en le marquant avec la souris dans la file — la sélection le met en inverse-vidéo — et en utilisant le bouton "**Éliminer**" (cf. Figures 5.1, 5.2). Par défaut, ce bouton est inactivé. Il n'est activé et donc utilisable que lorsqu'un document de la file de traitement est marqué.



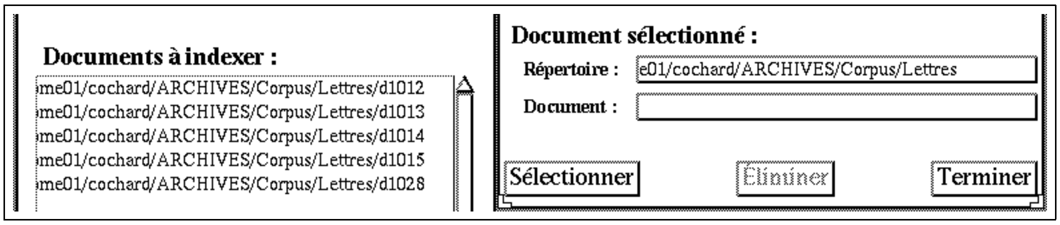


Figure 5 Lorsqu'un document est marqué dans la file de traitement, le bouton "**Éliminer**" est activé. L'exécution de "**Éliminer**" fait disparaître le document fautif de la file.

3.5 La terminaison de la sélection

Pour terminer une phase de sélection de documents, il est conseillé de faire disparaître la fenêtre de sélection en lançant la commande "**Terminer**". Le positionnement dans le système de fichiers est conservé entre la disparition et l'ouverture ultérieure de la fenêtre.

4 L'indexation et la consultation des résultats

Dans l'introduction nous avons présenté deux formes distinctes d'indexation : une indexation structurelle et une indexation linguistique. La première résulte d'une décomposition du document en blocs de texte et d'une identification de ces blocs. La deuxième utilise le résultat de la première pour isoler le bloc *sujet* et le traiter par des techniques d'analyse linguistique. Ces deux formes d'indexation sont groupées au sein de la commande "**Indexer**" qui fait subir au document une séquence de traitements et qui engendre les deux types d'index (cf. [Coc93a]).

4.1 La commande d'indexation et les messages d'exécution

Pour lancer le processus d'indexation, il est indispensable d'avoir constitué une file de traitement (cf. section 3). La commande "**Indexer**" traite un par un tous les documents de la file de traitement, en les éliminant au fur et à mesure[5]. Pour chaque document, une série de messages d'exécution est affichée; ce qui permet de suivre l'évolution du processus d'indexation et de patienter en connaissance de cause !

Il est possible que l'indexation ne se fasse pas complètement, c'est-à-dire jusqu'à la constitution d'un index linguistique, si bien que le traitement d'un document peut se terminer soit par le message "**a réussi !**", soit par le message "**a échoué !**". En cas d'échec, tous les traitements n'auront probablement pas été effectués. C'est le rôle de la commande "**Résultats...**" de consultation des résultats (cf. section 4.2) de présenter le résumé des opérations qui ont effectivement eu lieu sur chaque document.

Les différentes copies d'écran de la Figure 6 donnent la liste complète et chronologique des messages d'exécution produits durant le processus d'indexation. Chaque message est associé à un traitement précis dont le rôle est décrit ci-dessous :

Décomposition en blocs – cette étape rendue obligatoire par le format uniquement ASCII des documents à indexer, effectue un découpage géométrique des zones du texte en blocs;

Extraction des blocs significatifs – durant cette étape, l'ensemble des blocs est confronté à une grammaire de mise en page qui décrit différents modèles de rédaction de documents. Si le document candidat satisfait les règles de la grammaire, ses blocs reçoivent une étiquette et seuls certains d'entre eux sont retenus pour les besoins de l'indexation;

Filtre sur la taille du sujet – comme le prototype d'indexation linguistique qui traite le *sujet* du document est encore très sensible à la taille de la phrase à analyser, nous avons décidé de filtrer les sujets trop longs afin d'éviter des analyses trop longues, voire des crashs du système;

[5] Il ne s'agit pas d'une élimination physique du document; le logiciel ne touche absolument pas aux originaux. Il s'agit ici simplement de faire disparaître l'élément de la liste

Analyse syntaxique des composantes linguistiques – c'est une des deux étapes centrales, avec la suivante, du processus d'indexation. Actuellement, seuls les sujets contenant des mots connus sont analysés sans pour autant que le niveau de détail de l'analyse ne soit connu;

Génération d'une clé d'indexation – durant cette étape, quelques résultats d'analyse sont pris en compte — les plus vraisemblables — et les informations capitales pour l'indexation sont retenues et organisées afin de refléter certaines dépendances linguistiques profondes.

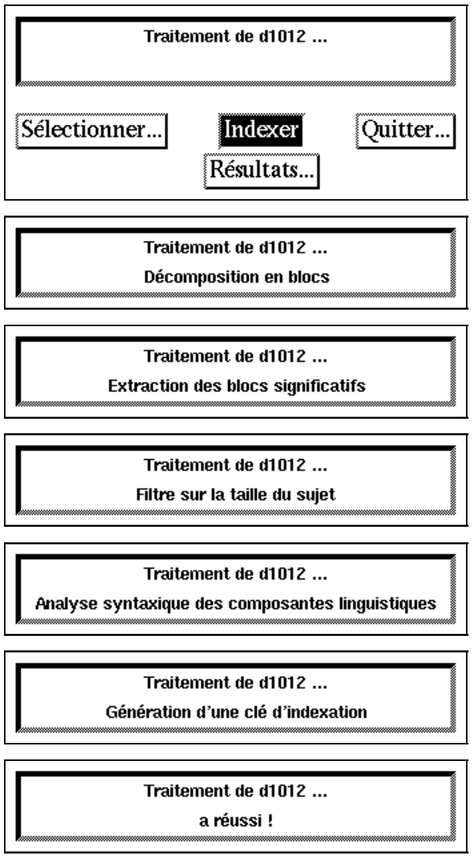


Figure 6 Lors du démarrage de l'indexation, le premier document de la file est pris en compte et un premier message est affiché. Les autres messages viennent compléter le premier en indiquant la phase d'indexation en cours. Le processus s'achève normalement après la création d'une clé d'indexation.

4.2 La consultation des résultats

La consultation des résultats d'indexation peut se faire après, ou durant, le processus d'indexation. La commande "**Résultats**..." ouvre une fenêtre dans laquelle s'inscrit, pour chaque document, une trace de son indexation.

La Figure 7 donne un exemple d'une telle trace avec un échantillon de tout ce qui peut arriver, ou presque ! durant une session de travail. Chaque ligne de la trace est produite par le traitement d'un document et contient les informations suivantes : le nom du document à indexer, le nom du document indexé entre parenthèses suivi du résumé des traitements subis. Chaque lettre correspond à une étape précise :

B – décomposition en blocs;

D – extraction des blocs significatifs (analyse du document);

H – filtre sur la taille du sujet (heuristique);

P – analyse syntaxique des composantes linguistiques ("parsing");

I – création d'une clé d'indexation.

Cette liste de caractères s'achève soit par "**ok**" qui indique que tout s'est effectué normalement, soit par "**##**" qui indique un échec lors du traitement qui précède ce symbole.

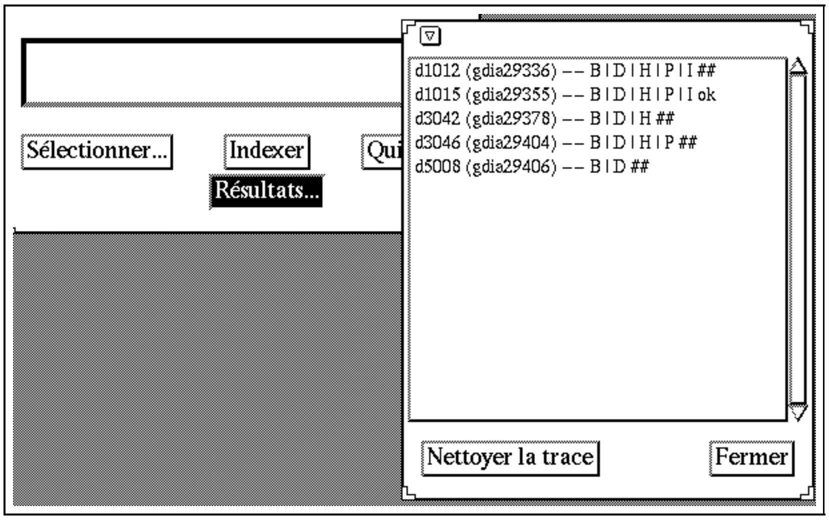


Figure 7 Un exemple d'une trace d'exécution offrant un échantillon des cas de figures possibles.

Afin de compléter l'information de l'utilisateur sur les résultats obtenus ou sur les raisons d'un échec, une fenêtre est affichée. Elle contient un complément d'information associé à l'entrée sélectionnée par la souris. Les Figures 8 à 12 présentent les commentaires associés à chacun des résultats présentés dans l'exemple de la Figure 7. Le bouton "**Fermer**" de cette fenêtre de commentaire (cf. Figure 8) permet de la faire disparaître.

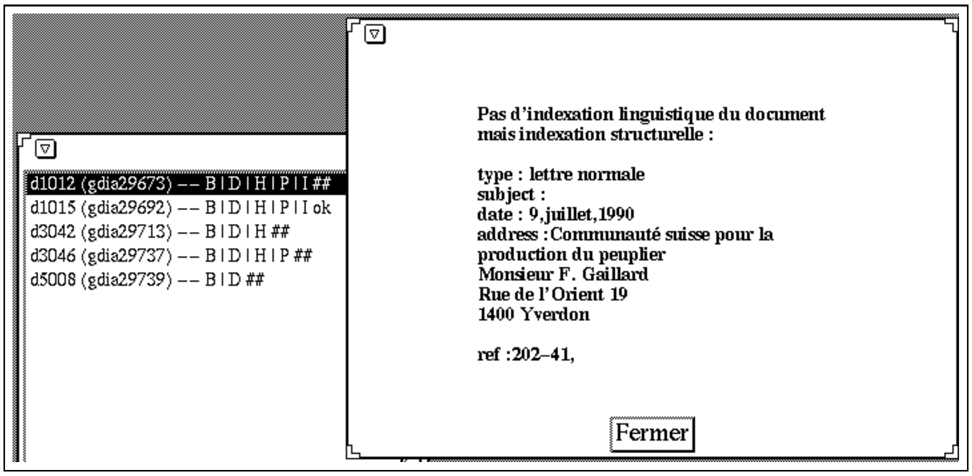


Figure 8 Cet exemple illustre un cas d'échec partiel d'indexation dû à l'absence d'un sujet dans la lettre. La fenêtre présente notamment les blocs qui ont été identifiés avec leur indentification, en particulier la date "**date :**", l'adresse du destinataire "**address :**", et le numéro de référence "**ref :**", le sujet "**subject :**" étant absent. Par ailleurs, chaque document est typé "**type :**". Seuls deux types sont reconnus actuellement : les *lettres normales* avec un destinataire précis et les *lettres circulaires*.

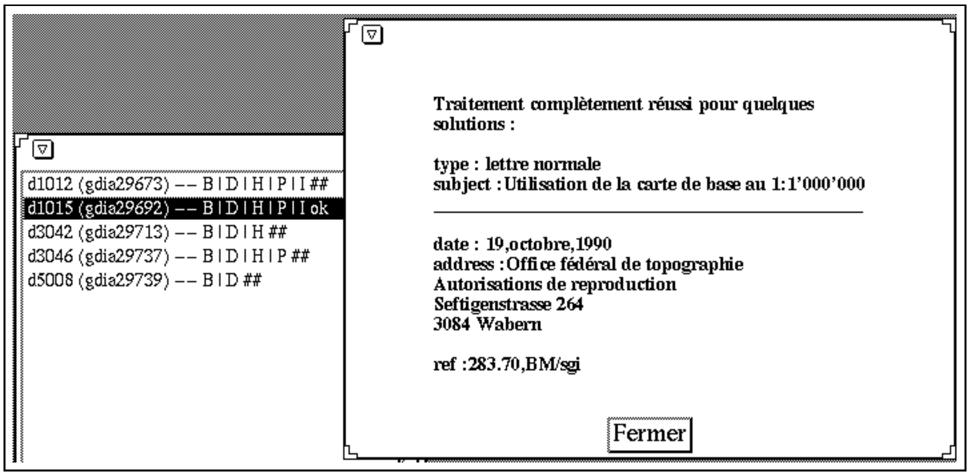


Figure 9 Cet exemple présente un cas d'indexation linguistique et structurelle complet. Les blocs identifiés sont affichés comme dans la Figure 8. Cette fois, un sujet était présent dans la lettre et il a ainsi pu donner lieu à un index linguistique. Cet index a été construit à partir de "**quelques solutions**", à savoir les plus vraisemblables. Deux autres stratégies sont possibles pour la construction d'un index : il peut s'agir d'un index construit à partir de "**la meilleure solution**" ou bien de "**toutes les solutions**". Pour l'instant, le choix d'une stratégie est déterminé uniquement par l'algorithme.

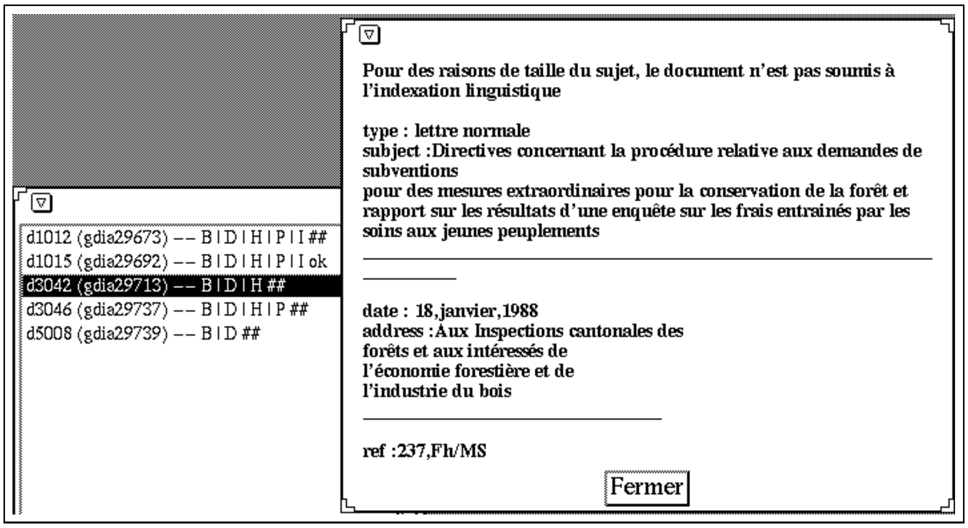


Figure 10 Cet exemple est typique d'une absence d'analyse linguistique causée par une taille trop grande du sujet. Actuellement le filtre rejette tous les sujets de plus de 120 caractères (seuil purement empirique et totalement provisoire). L'indexation structurelle est effectuée ici aussi.

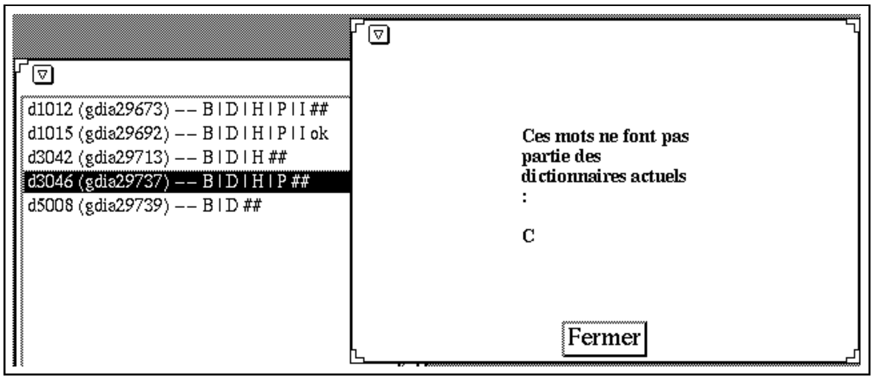


Figure 11 Dans cet exemple peu illustratif mais réel, le sujet contient la lettre "**C**" isolée (`Adaptation des formules A, C et ...`) qui ne fait pas partie de notre dictionnaire linguistique.

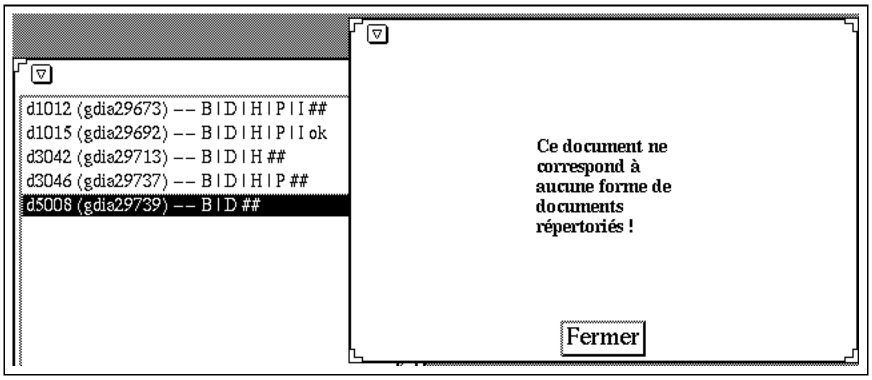


Figure 12 Ce dernier exemple illustre un cas d'analyse structurelle qui échoue. Ce document ne respecte pas les règles de mise en page qui ont été établies sur la base du corpus à disposition. Il s'agit soit d'une forme rare de mise en page, soit d'un bug possible bien que très improbable (!) de l'analyseur de structure.

Le bouton "**Fermer**" (cf. Figure 7) permet de faire disparaître la fenêtre d'affichage des résultats d'exécution et le bouton "**Nettoyer la trace**" (cf. Figure 7) permet, comme son nom l'indique, d'effacer le contenu de la trace d'exécution.

5 La terminaison du programme

La commande "**Terminer...**" du tableau de bord permet de mettre fin à une session de travail avec **I d'i 1.4**. Une confirmation est demandée avant de tuer tous les processus impliqués dans ce traitement (cf. Figure 13). La combinaison de touches `Control-C` est une autre solution pour mettre fin à une session. Elle donne lieu à la même demande de confirmation.

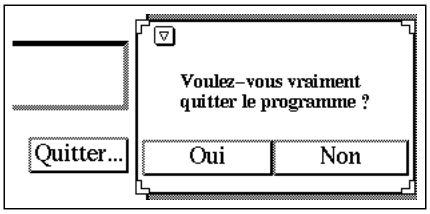


Figure 13 Fenêtre de confirmation de la commande de fin de session. Le dialogue est modal ce qui oblige l'utilisateur à répondre à la question posée avant de faire autre chose.

6 L'organisation logicielle

Pour pouvoir utiliser **I d'i 1.4**, il y a un certain nombre de prérequis plus ou moins simples à satisfaire :

- disposer d'une SPARCstation de SUN Microsystems;
- disposer du logiciel SICStus-Prolog 0.7 #9[6];
- disposer de l'interprète `wish` de la boîte à outils `Tk`[7] [Ous91];
- disposer de l'interprète `expect` (extension du language `Tcl`)[8] [Lib90];
- disposer des programmes `indexation.tcl`, `in_pro` et `ir`[9].

6.1 L'installation de I d'i 1.4 et la commande Unix

Pour installer l'interface d'indexation documentaire, il faut que les logiciels `prolog, expect et wish` soient accessibles directement, en complétant éventuellement le contenu de la variable d'environnement `PATH`.

Il faut aussi créer un répertoire dans lequel seront centralisés les index et les documents archivés. Ce répertoire doit être repéré par la variable d'environnement `ARCHIVEHOME` et doit avoir un sous-répertoire `Docs`. La série de commandes Unix suivante satisfait ces demandes :

```
% cd ~
% mkdir Archives
% mkdir Archives/Docs
% cat <<END >>.cshrc
setenv ARCHIVEHOME ${HOME}/Archives
END
```

La commande Unix qui démarre l'environnement démontré dans ce rapport est la suivante :

```
% in_pro indexation.tcl french ir >& /dev/null
```

Comme cet interface est bilingue, français-allemand, il est possible de démarrer l'environnement en précisant que les titres et messages doivent être fournis en allemand :

```
% in_pro indexation.tcl german ir >& /dev/null
```

Pour des raisons élémentaires de confort d'utilisation, il est conseillé de créer un “alias” plus concis.

6.2 Les opérations internes du lancement de I d'i 1.4

Le lancement de l'exécution de **I d'i 1.4** est une opération assez longue. Quelques explications sur le chargement et l'interaction des processus de ce système seront utiles pour donner un sens à la commande Unix décrite ci-dessus. En outre, cela permettra à l'utilisateur de patienter en connaissance de cause.

nécessaires au fonctionnement de ce système.

Le programme qui gère le démarrage de tout ce système est `in_pro` qui a deux rôles : le premier est de lancer l'exécution des programmes qui figurent en paramètres, à savoir `indexation.tcl french` et `ir`; le deuxième est d'assurer la communication entre ces deux programmes “fils”.

L'interface défini dans `indexation.tcl` est un programme de taille modeste qui est assez rapidement chargé. Par contre `ir` est un très gros programme — sa taille actuelle est supérieure à 10 MB — et son chargement prend du temps.

[6] Adresse électronique à contacter : `sicstus_request@sics.se`.
[7] Site pour faire un “ftp” : `barkley.berbeley.edu`.
[8] Site pour faire un “ftp” : `ftp.cme.nist.gov`.
[9] Adresse électronique à contacter : `cochard@idiap.ch`.

6.3 Les fichiers de langues

L'interface **I d'i 1.4** est bilingue, français-allemand. Les noms, titres et messages dans chacune des deux langues sont centralisés dans deux fichiers paramètres : `.IndexationFrench` et `.IndexationGerman`. Il s'agit de fichiers texte qui peuvent être édités et modifiés à la convenance de l'utilisateur. Ces fichiers sont des fichiers de ressources dans la terminologie X [QO90] et toute adaptation du contenu doit respecter les règles des fichiers de ressources de X. Pour que **I d'i 1.4** trouve ces fichiers paramètres, il est indispensable que chaque utilisateur en ait une copie dans son répertoire `$HOME`.

Le fichier paramètre pour le français a le contenu suivant :

```
*Button.font: *Times-medium-r-*-18-*
*Button.borderwidth: 3

*select_b.text: Sélectionner...
*index_b.text: Indexer
*quit_b.text: Quitter...
*show_log_b.text: Résultats...
*select_action_b.text: Sélectionner
*cut_action_b.text: Éliminer
*quit_action_b.text:  Terminer
*yes_b.text: Oui
*no_b.text: Non
*close_b.text: Fermer
*clear_b.text: Nettoyer la trace

*main_title.text: Indexation documentaire
*list_title.text: Documents à indexer :

*dirs_title.text: Répertoires :
*files_title.text: Documents :

*selection_title.text:  Document sélectionné :
*dir_label.text: Répertoire :
*file_label.text: Document :

*quit_mess.text: Voulez-vous vraiment quitter le programme ?

*mess.treatment.text:  Traitement de
*mess.blocs.text: Décomposition en blocs
*mess.document.text: Extraction des blocs significatifs
*mess.heuristic.text: Filtre sur la taille du sujet
*mess.parsing.text: Analyse syntaxique des composantes linguistiques
*mess.indexing.text: Génération d'une clé d'indexation
*mess.best.text: pour la meilleure solution :
*mess.some.text: pour quelques solutions :
*mess.all.text: pour toutes les solutions :
*mess.succeeded.text: a réussi !
*mess.failed.text: a échoué !
*mess.initialisation.text: Chargement du système
*mess.done.text: terminé
*mess.fichier_absent.text: Pas de fichier à sélectionner

*mess.b_err1.text:  Décomposition impossible du document en blocs. \
```

```
Comportement très étrange !
*mess.b_err2.text: Document inexistant !
*mess.d_err1.text: Ce document ne correspond à aucune forme de documents \
répertoriés !
*mess.h_err1.text:      Pour des raisons de taille du sujet, le document \
n'est pas soumis à l'indexation linguistique
*mess.p_err1.text: Ces mots ne font pas partie des dictionnaires actuels :
*mess.i_err1.text: Pas d'indexation linguistique du document mais \
indexation structurelle :
*mess.i_res.text: Traitement complètement réussi
```

7 Commentaires et conclusion

I d'i 1.4 a été développé sur des SPARCstation SUN et fonctionne dans l'environnement OpenWindows de SUN. L'interface graphique a été réalisé à l'aide de la boîte à outils graphique Tk; l'indexation est prise en charge par un programme écrit en SICStus-Prolog 0.7 #9 et la communication entre ces deux applications est réalisée à l'aide d'une extension de Tcl, appelée Expect.

Cette application en collaboration avec **I d'a 1.0** et **I de r 1.0** préfigure ce que pourrait être un environnement d'indexation automatique et de recherche documentaire dans un cadre de travail comme les Archives fédérales suisses ou tout autre institution qui manipule de grosses bases de données textuelles.

7.1 Problèmes connus de l'interface

Dans sa version actuelle, **I d'i 1.4** souffre d'un certain nombre de problèmes de jeunesse qui seront corrigés prochainement. Voici, pour information, une liste de problèmes déjà recensés :

- Lorsqu'on fait un accès direct sur un répertoire qui n'existe pas, dans la fenêtre de sélection des documents, le système engendre un message d'erreur parasite dans la liste des documents.
- Lors d'une tentative d'indexation d'un document inexistant, le système se bloque.
- Absence de titres aux fenêtres car le bilinguisme devrait s'appliquer ici aussi.
- Absence d'options de fonctionnement qui permettrait, par exemple, de déterminer le niveau de création des index.
- Création d'une erreur interne non fatale lorsqu'on tente de consulter un commentaire sur une trace d'exécution qui n'est pas complète.
- Un même document peut être indexé plusieurs fois sans que le système ne réagisse.

7.2 Amélioration de l'analyseur linguistique

L'utilisation de l'analyseur linguistique (le programme `ir`) laisse entrevoir quelques faiblesses qu'il est indispensable de corriger avant d'envisager d'étendre ses dictionnaires ou sa couverture grammaticale.

- Concernant le traitement purement linguistique, l'heuristique devrait disparaître au profit d'un analyseur moins sensible à la taille des textes qu'on lui soumet.
- La stratégie d'analyse doit être modifiée pour permettre à la fois d'effectuer une analyse partielle même en présence de mots inconnus.

Toute suggestion d'amélioration ou compte rendu de problèmes à l'adresse électronique figurant dans l'entête de ce document sont les bien venus.

Références

[CHK93] Jean-Luc Cochard, Michael Hess, and Andreas Kellerhals. A linguistically based information retrieval system for administrative letters. Technical report TR-93-xx, IDIAP, Institut Dalle Molle d'Intelligence Artificielle Perceptive, 1993.

[Coc93a] Jean-Luc Cochard. Une interface d'administration de documents index s, i d'a 1.0. Rapport technique RT-93-02, IDIAP, Institut Dalle Molle d'Intelligence Artificielle Perceptive, janvier 1993.

[Coc93b] Jean-Luc Cochard. Une interface de recherche documentaire, i de r 1.0. Rapport technique RT-93-03, IDIAP, Institut Dalle Molle d'Intelligence Artificielle Perceptive, janvier 1993.

[Lib90] Don Libes. The expect user manual – programmatic dialogue with interactive programs. Nist ir, National Institute of Standards and Technology, November 1990.

[Ous91] John K. Ousterhout. An X11 toolkit based on the Tcl language. In *Proceedings of the 1991 Winter USENIX Conference*, 1991.

[QO90] Valerie Quercia and Tim O'Reilly. *X Window System User's Guide for X11 R5*. O'Reilly and Associates, Inc., 1990.