

PAVLIDES Marina 0698

DESS  
1984  
10  
B

# RAPPORT DE STAGE

**RHÔNE-POULENC AGROCHIMIE**



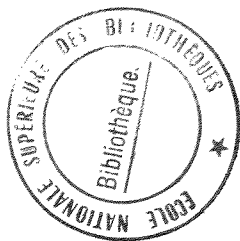
UNIVERSITE CLAUDE BERNARD LYON I

D.E.S.S. INFORMATIQUE DOCUMENTAIRE

RAPPORT DE STAGE :

MEMORISATION ET TRAITEMENT

DE REFERENCES BREVETS DE CHIMIE



PAVLIDES Marina

JUIN 1984

**RHÔNE-POULENC AGROCHIMIE**

14-20, rue Pierre Balzet Lyon 9<sup>e</sup> - B.P. 9163 - 69263 Lyon Cedex 1 - Tél. (7) 864.81.81 - Téléx: rhône 310098 F

Je remercie M. Brachotte, responsable du Service Propriété Industrielle et Documentation, de m'avoir accepté en stage, et M. Faraut, responsable du Service Documentation.

Je remercie également M. Rubio, ingénieur au Service Informatique, pour avoir bien voulu superviser mon travail et assister à la soutenance de ce rapport.

S'il ne m'est pas possible de citer toutes les personnes qui m'ont aidée, je tiens à les remercier de leur accueil et de leurs conseils.

Je remercie tout particulièrement M. Marmain, ainsi que M. Cléménçon du Centre Rhône-Poulenc St Fons.

## P L A N

### I/ INTRODUCTION

### II/ GENERALITES

- 1 - La Société Rhône-Poulenc Agrochimie
- 2 - Le Centre de Recherche de La Dargoire
- 3 - Le Service Informatique
- 4 - Le Service Propriété Industrielle et Documentation

### III/ RECHERCHE EN LIGNE DE BREVETS DE CHIMIE

- 1 - Bases de données interrogées
- 2 - But et nature des recherches de références brevets
- 3 - Sujet du stage

### IV/ RAPPEL SUR LES BREVETS

- 1 - Etablissement d'un brevet
- 2 - Numérotation des brevets

### V/ COMPARAISON DES DEUX BASES DE DONNEES

- 1 - Généralités
- 2 - Stratégies de recherche
- 3 - Références brevets obtenues
- 4 - Format des numéros de brevet

### VI/ CHOIX D'UNE APPLICATION

- 1 - Interrogations. Emission et réception des données
- 2 - Fichier résultat
- 3 - Programmation
- 4 - Enchaînement des travaux

### VII/ PROGRAMMES DE TRAITEMENT DE TEXTE

- 1 - Lexique n° 1
- 2 - Lexique n° 2
- 3 - Lexique n° 3

### VIII/ FICHER RESULTAT

### IX/ CONCLUSION

### ANNEXES

## I / INTRODUCTION

Ce stage, d'une durée de 4 mois (du 5 mars au 29 juin), a été effectué dans la Société Rhône-Poulenc Agrochimie, au Service Propriété Industrielle et Documentation du Centre de Recherche de La Dargoire à Lyon, sous la direction de M. Brachotte.

En vue d'élaborer de nouveaux axes de recherche pour les chercheurs de la Société, le Service Documentation interroge, entre autres, des bases de données bibliographiques externes dans le domaine des Brevets de chimie.

Le sujet de ce stage consistait à choisir et mettre en place une application permettant :

- d'interroger la base "Derwent Patents" par émission automatique d'une liste de numéros de brevets provenant des références brevets de la base "Chemical Abstracts" précédemment interrogée ;

- de réunir les références brevets obtenues par l'interrogation des deux bases de données, de manière à les présenter en une liste unique et homogène, à l'information sélectionnée, fusionnant les références traitant des mêmes brevets.

Cette application a été réalisée par interrogation des bases de données par terminal relié au miniordinateur du Service Informatique, traitement des références mémorisées par des programmes de traitement de texte, création d'un fichier Texto pouvant permettre par la suite, d'étudier les références brevets enregistrées et de les éditer selon des critères de sélection variés.

## II/ GENERALITES

### 1 - LA SOCIETE RHONE-POULENC AGROCHIMIE

Rhône-Poulenc Agrochimie, l'une des Divisions de la société nationalisée Rhône-Poulenc, est l'un des leaders dans le domaine de la protection des cultures.

12 filiales sont implantées à l'étranger, principalement en Europe de l'Ouest, en Amérique du Nord et au Brésil.

L'un des secteurs essentiels de la Société Rhône-Poulenc Agrochimie est constitué par la Recherche et le Développement :

Il existe 3 Centres de recherche : Paris-Vitry, Lyon-La Dargoire, Dagenham-Ongar (en Grande-Bretagne), où travaillent environ 700 personnes, dont 200 ingénieurs et cadres. Ces Centres sont en liaison avec 6 Stations expérimentales fixes (3 en France, 1 en Grande-Bretagne, 1 au Brésil), et des installations mobiles, principalement aux USA, au Soudan, en Côte-d'Ivoire.

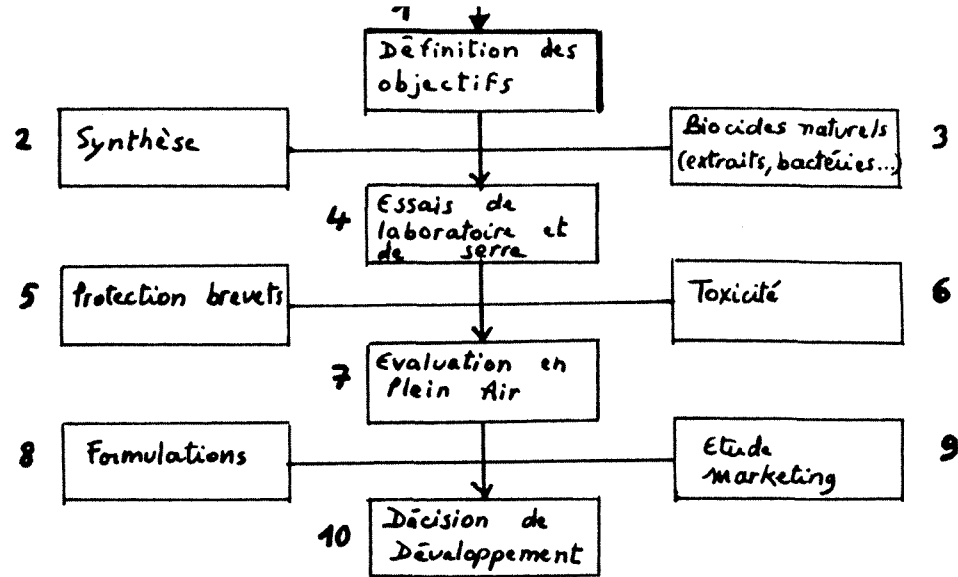
La première activité industrielle de Rhône-Poulenc Agrochimie concerne la fabrication de matières actives originales (résultant des recherches) ainsi que des matières actives classiques ; le contrôle qualitatif de ces produits est particulièrement important.

Par ailleurs la mise au point de nouvelles formulations des produits (micro-encapsulation,...) représente l'autre aspect de cette activité industrielle.

L'organigramme de la Société Rhône-Poulenc Agrochimie est donnée page suivante.

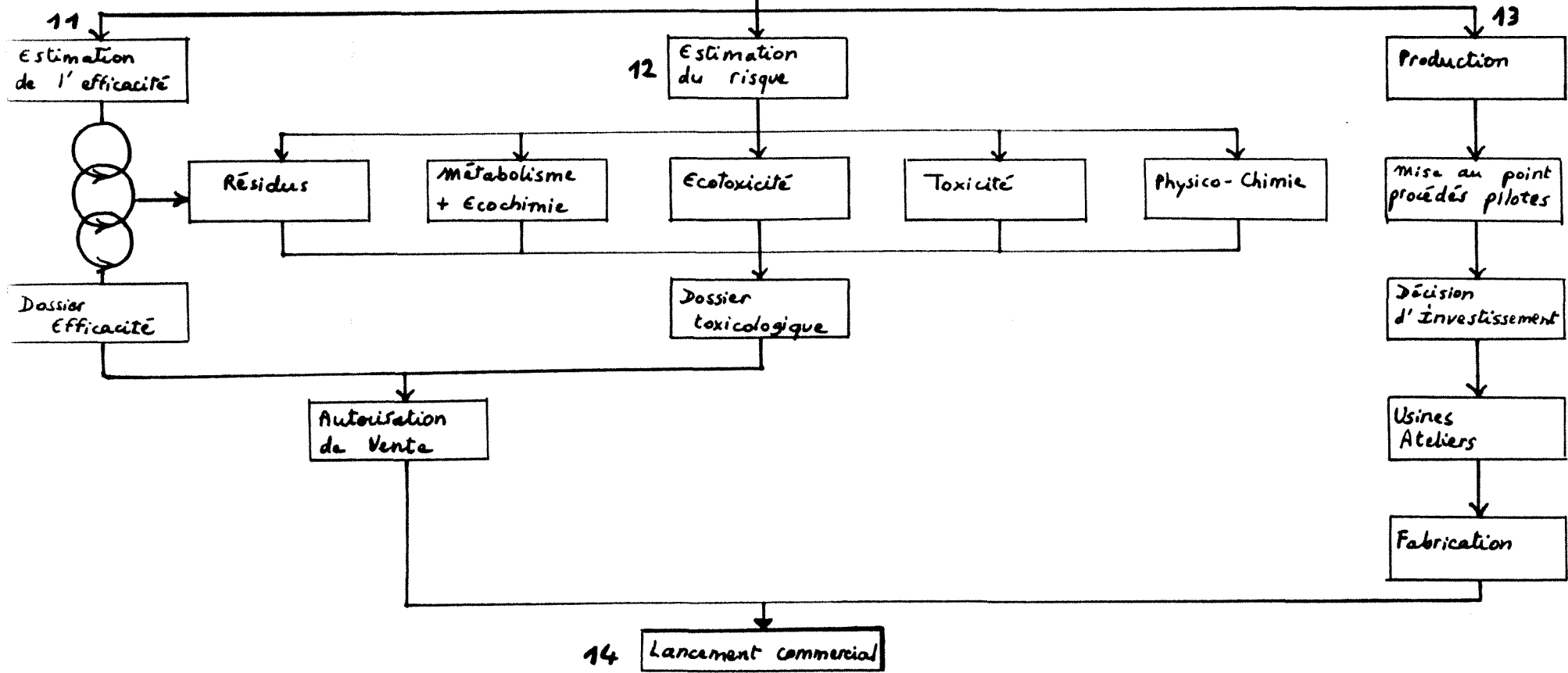
Les étapes successives aboutissant à la sélection d'une molécule active parmi des milliers et au lancement commercial du produit correspondant, sont ensuite schématisées.

La nécessité d'une information complète et pertinente (activités documentaires) apparaît au niveau de plusieurs de ces étapes, en particulier au niveau de la première étape, essentielle : la définition des objectifs.



Les étapes successives aboutissant au lancement commercial d'un produit

(= Sélection d'une molécule parmi des milliers d'autres)



## 2 - LE CENTRE DE RECHERCHE DE LA DARGOIRE

Situé dans le neuvième arrondissement de Lyon, le Centre de Recherche de La Dargoire a été créé en 1966 pour répondre aux besoins de l'expansion de la Société Pechiney-Progil ~~intégrée~~ par la suite le groupe Rhône-Poulenc.

En 1977 le Centre de La Dargoire est devenu le siège social de la Société Rhône-Poulenc Agrochimie, filiale de Rhône-Poulenc.

Depuis le Centre poursuit son extension ; c'est ainsi que de nouvelles serres ont été construites à proximité des laboratoires de recherche ; actuellement des travaux d'agrandissement sont en cours pour le réaménagement des différents services.

Le Centre de La Dargoire, où travaillent 428 personnes (dont 131 cadres et 203 agents de maîtrise et techniciens) est constitué de différents Services.

Suivent une présentation plus précise du Service Informatique, et plus particulièrement une présentation du Service Propriété Industrielle et Documentation, où ce stage a été effectué.



Centre de Recherche - Lyon/La Dargoire



### 3 - LE SERVICE INFORMATIQUE

#### 3-1 Télécommunications

Il existe pour l'ensemble du groupe Rhône-Poulenc, un Service Informatique général avec deux centres de traitement :

- le Centre de Courbevoie
- le Centre du Tonkin, à Villeurbanne.

Ces deux centres où l'information est traitée par des ordinateurs puissants (IBM ...) sont reliés par un réseau spécialisé à grande vitesse.

Les télécommunications par réseaux spécialisés en provenance ou à destination du Centre de Recherche de La Dargoire se font généralement par l'intermédiaire du Centre du Tonkin (la liaison entre La Dargoire et le Tonkin étant réalisée par ligne commutée téléphonique classique).

#### 3-2 Le Service Informatique au Centre de La Dargoire

Le Service Informatique a des rôles multiples, dont la coordination, la maintenance et le développement des applications informatiques dans le Centre.

L'informatique est utilisée à des fins variées et avec plus ou moins d'ampleur, par divers Services : Comptabilité, Achats, Développement, Recherche, Documentation ...

Ainsi, en recherche, il existe plusieurs applications informatiques, telles que la synthèse assistée par ordinateur, les traitements statistiques d'analyse des données expérimentales,...

#### 3-3 Matériel informatique en service à La Dargoire

Le Centre de recherche utilise, outre divers terminaux, minitels, et microordinateurs, le système WANG OIS :

26 périphériques (microordinateurs orientés vers le traitement de texte, unités d'archivage à distance, imprimantes, poste de télécommunication), répartis dans les différents services du Centre, sont reliés à l'unité centrale (la capacité maximum du système étant de 32 périphériques).

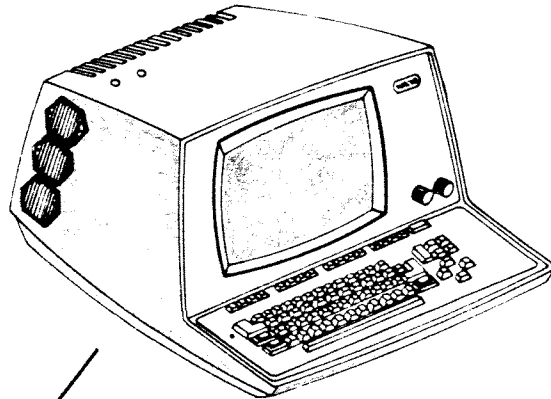
La mémoire de masse du disque-système (partagée entre les divers Services) se compose de 5 disques durs, représentant 64 millions de caractères (64 mégaoctets). Par ailleurs, un disque amovible de 13 mégaoctets assure la sauvegarde du système.

Chaque machine de traitement de texte est un microordinateur possédant sa mémoire propre, (un "terminal intelligent") en même temps qu'un terminal à usages multiples. En ce qui concerne le traitement de texte, il est possible de créer des programmes particuliers (lexiques), ou de gérer des fichiers (fichiers d'adresses,...) avec l'option "gestion de liste".

D'autre part, les Documents de traitement de texte (information enregistrée sur le Disque-système) peuvent être archivés (sauvegardés), au moyen d'unités d'archivage à distance, sur diquette 8 pouces (1200 secteurs de 256 caractères). Des sauvegardes quotidiennes sont ainsi effectuées, et tous les documents importants sont conservés en archive.

Les principales applications réalisables avec le système Wang OIS sont résumées sur le schéma de la page suivante.

PRINCIPALES APPLICATIONS  
du système WANG O.I.S.



Terminal  
de Traitement de Texte  
WANG OIS

relié à l'ordinateur central  
possède une mémoire interne

↓  
Traitement de texte  
création de document  
révision  
impression  
archivage sur disquette  
...

↓  
Index des documents  
- sur le disque-système  
- sur une disquette  
- en impression ou  
attente d'impression

↓  
Utilitaires :  
reconstruction de document  
effacement de document  
tri de document  
duplication de document  
...

↓  
Gestion de listes  
(fichiers à champs fixes)

↓  
Applications BASIC

↓  
Messages

↓  
fusion de documents

↓  
Télécommunications

lancement d'une session "batch"  
lancement d'une session interactive  
...

↓  
Lexiques

(programmes de traitement de texte)

...

#### 4 - LE SERVICE PROPRIÉTÉ INDUSTRIELLE ET DOCUMENTATION

L'ensemble du Service Propriété Industrielle et Documentation représente un effectif de 20 personnes.

##### 4-1 Le Service Propriété Industrielle

Dans un secteur tel que l'Agrochimie, la Propriété Industrielle joue un rôle essentiel.

Le Service Propriété Industrielle, a pour rôle la protection des inventions et découvertes de la Société Rhône-Poulenc Agrochimie, par l'obtention de Brevets en France à l'I.N.P.I. (Institut National de la Propriété Industrielle), et à l'étranger dans certains pays (selon la nature des découvertes) par l'intermédiaire de l'Office Européen des Brevets et d'Agents de brevets.

Il peut s'agir de la protection d'un produit chimique, d'une composition, d'un procédé de traitement, d'un procédé de préparation...

Lorsque l'invention est jugée brevetable, le Service délimite de manière très précise et en collaboration avec les chercheurs, le contenu des revendications de la Demande de Brevet (c'est-à-dire l'objet de la protection). La délimitation des revendications doit être très précise et chaque mot est important.

Le Service Propriété Industrielle doit également maintenir en vigueur les Brevets par le paiement d'annuités, et agir contre d'éventuelles contrefaçons.

Un fichier "Portefeuille Brevets" (logiciel Texto) recense tous les brevets déposés par Rhône-Poulenc Agrochimie (en fait plusieurs fichiers sont constitués).

##### 4-2 Le Service Documentation

La bibliothèque est constituée de quelques 3000 ouvrages, de 400 périodiques spécialisés traitant de chimie, d'agronomie, d'économie, de toxicologie et d'environnement,...

On peut citer la collection complète des chemicals abstracts manuels (résumés et index divers).

Le Service Documentation gère des banques de données internes ; ces fichiers, qui utilisent le logiciel Texto de la Société Chemdata, sont constitués soit à partir des publications reçues (fichier de documentation agronomique, fichier de documentation économique...), soit à partir des rapports de recherche et des rapports expérimentaux.

Ils sont interrogeables à distance par les filiales de Rhône-Poulenc Agrochimie, (telles que "May and Baker" en Grande-Bretagne), ainsi que par les autres succursales du groupe Rhône-Poulenc (collaboration avec le secteur pharmaceutique, en particulier)

A partir des revues spécialisées sont constitués des bulletins d'information qui sont diffusés à l'ensemble des Services

("Agridoc": actualités dans le domaine agronomique, et "Docéco" : nouvelles économiques).

Par ailleurs des profils de diffusion interne sont réalisés pour les personnes nécessitant une information très spécialisée (par exemple, un chercheur souhaitant être tenu au courant de toute information sur un produit donné) par l'interrogation des fichiers internes (logiciel Texto) et des bases de données externes.

Certains abonnements à des revues spécialisées sont également personnalisés.

Les banques de données externes sont fréquemment interrogées, soit pour la recherche ( Chemical Abstracts : CAS online, Biosis, Derwent : World Patents Index, Agridoc...), soit pour les études de toxicologie (Toxline, Medline, Derwent Pestdoc...), soit pour le Service Propriété Industrielle (INPI,...)...

Sur le plan informatique, le Service Propriété Industrielle et Documentation dispose de :

- Système Wang OIS :

Les périphériques suivants sont utilisés :

- 3 machines de traitement de texte, dont 2 reliées à une unité d'archivage, et 1 reliée à un modem (donc pouvant servir de terminal d'interrogation)
- 1 imprimante feuille à feuille (20 car./s).
- 2 Minitels
- 5 terminaux d'interrogation (Tektronix 4012, Texas Instrument Silent 700,...).

En conclusion, on peut noter que le Service Propriété Industrielle et Documentation du Centre de la Dargoire, Service important au sein de la Société Rhône-Poulenc Agrochimie et traitant une information considérable et de nature variée, utilise depuis de nombreuses années, les possibilités de l'informatique documentaire.

### III/ RECHERCHE EN LIGNE DE BREVETS DE CHIMIE

#### 1 - BASES DE DONNEES INTERROGEEES

Le Service Propriété Industrielle et Documentation de La Dargoire interroge fréquemment, pour des recherches de brevets de chimie, les bases de données suivantes :

- Base américaine "Chemical Abstracts" (fichiers Registry et CA) par le serveur CAS ONLINE (le producteur est son propre serveur, avec un logiciel donc bien adapté à la base de données) ;
- Base anglaise "Derwent Patents" (fichiers WPI et WPIL), par l'intermédiaire du serveur américain SDC (logiciel d'interrogation ORBIT).

Les autres bases de données bibliographiques donnant des références de brevets (bases Claims, ,...) ne sont pas interrogées. En matière de propriété industrielle, les bases de données de l'I.N.P.I. (Institut National de la Propriété Industrielle) sont parfois interrogées.

#### 2 - BUT ET NATURE DES RECHERCHES EFFECTUEES

Les recherches de brevets pratiquées par l'interrogation en ligne des bases de données Chemical Abstracts et Derwent Patents, répondent à différentes préoccupations :

- Recherches documentaires d'antériorité, lorsque sont découvertes des molécules aux propriétés intéressantes, ou des effets synergiques résultant de l'association de matières actives, par exemple :

Ces interrogations ont pour but de vérifier l'originalité de ces découvertes, en recherchant si des brevets n'ont pas déjà été déposés par d'autres sociétés, couvrant ces résultats.

Dans le cas contraire, il y a innovation et Rhône-Poulenc peut alors protéger ces découvertes par le dépôt de brevets.

Ce genre de recherches documentaires, sur des molécules précises, conduit généralement à un nombre faible, voire nul en cas d'innovation, de références brevets pertinentes.

- Recherches de "documentation stratégique", de tous les brevets ayant trait à une famille ou sous-famille de composés :

Ces recherches permettent de cerner les résultats déjà obtenus dans tel domaine, les activités de recherche et les intérêts des laboratoires concurrents, et de déterminer ainsi des axes de recherche, en partie donc en fonction des brevets déjà déposés.

Ces recherches documentaires, actuellement peu fréquentes (environ une à deux fois par mois), sont larges, et fournissent des listes plus ou moins importantes de références brevets (plusieurs dizaines à plus d'une centaine) selon les groupes de molécules étudiés.

Dans tous les cas, les deux bases de données sont en général interrogées, par souci d'exhaustivité. En effet, ces deux bases ne recouvrent pas exactement les mêmes documents primaires (brevets) (les brevets de certains pays sont enregistrés dans la base Derwent mais pas dans la base Chemical Abstracts, par exemple), et par ailleurs elles impliquent des stratégies de recherche différentes.

Le plus souvent la base Chemical Abstracts est d'abord interrogée, puis la base Derwent Patents, dont les références, plus complètes, permettent, par le numéro d'accès attribué par Derwent, de retrouver le texte intégral des brevets intéressants (le Service Propriété Industrielle et Documentation possédant la collection des textes intégraux de brevets publiée par Derwent).

Traitement des listes de brevet obtenues par documentation stratégique :

Lorsqu'on effectue une recherche d'investigation assez large, les deux listes de brevets obtenues par l'interrogation des deux bases de données, constituent une information importante, et difficile à utiliser sans un dépouillement fastidieux.

En effet, les deux listes possèdent des références communes ; les références ne fournissent pas exactement les mêmes renseignements ; elles sont présentées différemment, et non classées.

Un autre inconvénient important vient du fait que l'interrogation de chacune des bases fait appel à un logiciel d'interrogation différent, et surtout impose une stratégie de recherche différente, les deux bases n'offrant pas les mêmes possibilités d'interrogation (interrogation par un numéro caractéristique du composé, ou par codes des structures, ou par dessin de la formule développée de la molécule,...).

### 3 - SUJET DE STAGE

Le but de ce stage était donc le suivant :

Etudier les possibilités et les moyens pratiques à mettre en oeuvre, pour :

- interroger les bases de données à l'aide d'un terminal relié au miniordinateur Wang OIS du Service Informatique ;

- enregistrer dans un Document traitement de texte (sur le disque-système) les références brevets obtenues par l'interrogation de la base Chemical Abstracts sur une famille de molécules ;

- Sélectionner à l'aide d'un programme, les numéros de brevet dans ces références et interroger la base Derwent en émettant automatiquement vers cette base, la liste de ces numéros de brevet sous forme d'étape de recherche, selon le logiciel d'interrogation Orbit du serveur SDC ;

- mémoriser dans un Document les références ainsi obtenues par cette deuxième interrogation ;

- traiter automatiquement l'ensemble des références obtenues, de manière à constituer une liste homogène de références brevets, classée, contenant une information sélectionnée et sans doubles, sous une présentation constante.

#### IV/ RAPPEL SUR LES BREVETS

Le Brevet d'Invention, délivré en échange de la divulgation d'une invention ou d'une découverte, accorde à son titulaire un monopole d'exploitation, limité dans le temps, et limité géographiquement (brevets par pays; brevet européen; brevet "international" PCT couvrant plus de pays que le brevet européen).  
En France ce monopole dure généralement 20 ans.

##### 1 - ETABLISSEMENT D'UN BREVET

Lorsqu'une première demande de brevet est déposée dans le pays d'origine, on attribue un numéro de dépôt. La date de dépôt est aussi la date de première priorité (anglais: "application date").

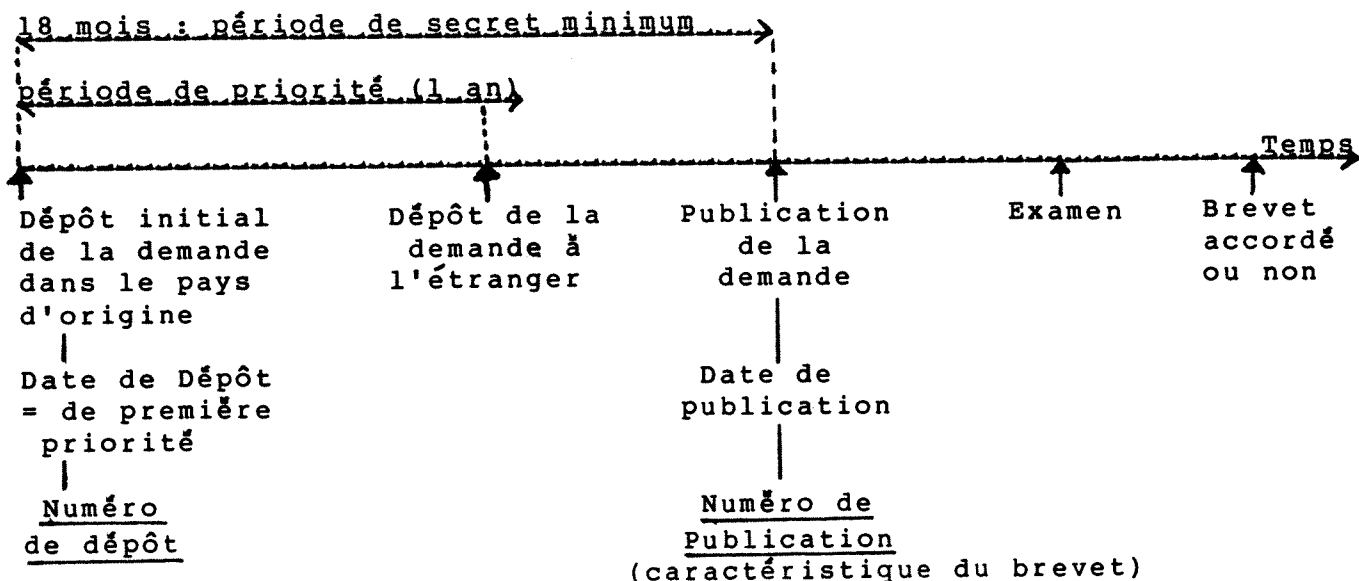
D'autres demandes de brevet sont ensuite généralement déposées à l'étranger. On peut faire également une demande de brevet européen, ou une demande de brevet international.

Une fois les demandes de priorité et de secret expirées, le texte du brevet est publié (l'invention est donc divulguée) et un numéro de publication est attribué. Le brevet est ensuite éventuellement accordé. Le numéro de publication attribué par le pays où le brevet a été déposé et publié est caractéristique du brevet.

En France les demandes de brevets sont déposées à l'I.N.P.I. (Institut National de la Propriété Industrielle).

Remarque : Pour certains pays (Danemark, Norvège, Finlande, Autriche,...), il n'existe pas de numéro de publication proprement dit, mais un numéro d'accord, lorsque le brevet est accordé.

Etapes successives conduisant à l'établissement d'un brevet :  
(cas général) :



Il existe de nombreuses catégories de brevets, et qui varient selon les pays. Ainsi, en France, existent :

- Les Brevets d'Invention
- Les Certificats d'Utilité
- Les Certificats d'Addition à un Brevet d'Invention
- Les Certificats d'Addition à un Certificat d'Utilité

On peut noter que les brevets constituent une information technique et scientifique considérable ( plus de 600 000 brevets sont déposés dans le monde chaque année ), mais encore sous-utilisée souvent (par rapport aux articles de périodiques et aux rapports de recherche ).

## 2 - NUMEROTATION DES BREVETS

En ce qui concerne les numéros caractéristiques des brevets, chaque pays a ses propres règles de numérotation.

Un numéro de dépôt est normalement attribué lors du dépôt des demandes de brevet. La numérotation de ces demandes est généralement recommencée chaque année à partir du numéro 1, de sorte que les deux derniers chiffres de l'année sont intégrés au numéro de dépôt, pour éviter toute confusion.

La plupart des pays attribuent ensuite un numéro de publication; pour ces numéros de publication tous ces pays numérotent en continu les brevets, d'année en année (exception : le Japon, qui intègre donc l'année dans le numéro de publication). A ce numéro de publication est associé le code normalisé international ICIREPAT de l'I.S.O. pour la représentation des noms de pays. Chaque pays est codé par deux lettres qui lui sont caractéristiques. (les Brevets européens sont codés EP et les Brevets PCT internationaux, WO).

Le numéro de publication et le numéro de dépôt peuvent différer.

On trouvera en annexe, la "recommandation sur la numérotation des documents de brevets publiés" publiée par l'O.M.P.I.



## V/ COMPARAISON DES DEUX BASES DE DONNEES

### 1 - GENERALITES

#### 1-1 Base Chemical Abstracts

La base de données de CAS Online comporte deux fichiers d'interrogation :

##### - Fichier "Registry" :

Il contient environ 6,5 millions de structures chimiques (molécules ou radicaux). Sa mise à jour est hebdomadaire, et plus de 6000 nouvelles substances sont enregistrées chaque semaine.

Ce fichier est interrogeable, en particulier, par les structures des molécules. Il ne fournit pas de références bibliographiques, celles-ci sont obtenues par sauvegarde de la recherche sur le fichier CA où ces références sont alors demandées.

##### - Fichier "CA" :

Ce fichier bibliographique contient à ce jour 6,2 millions de références, et couvre toute la littérature en chimie : brevets, mais aussi articles de périodiques, rapports, thèses ... Il est possible de sélectionner les brevets (par une recherche sur champ "type de document").

Environ 960000 brevets sont contenues dans le fichier CA. La couverture de la base en matière de brevets, est donnée en annexe.

#### 1-2 Base Derwent Patents

L'information primaire à l'origine de cette base (World Patents Index = WPI) est constituée exclusivement de brevets de chimie. C'est l'une des bases de données les plus utilisées dans le monde, et elle recense les brevets de plus de 50 pays, soit plus de 3 millions de familles de brevet. (Par ailleurs les Publications Derwent constituent le premier éditeur et producteur privé d'informations en matière de brevets).

Il existe plusieurs sections (fichiers d'interrogations indépendants) selon les domaines de la chimie concernés :

Domaines pharmaceutique (depuis 1963), Chimie agronomique (depuis 1965), Plastiques (depuis 1966), Chimie (depuis 1970), ...

En général la recherche de brevets est effectuée sans sélectionner l'une de ces sections, afin d'avoir la meilleure exhaustivité possible.

La base World Patents Index est décomposée en deux fichiers distincts :

- fichier WPI constitué par des références de brevets enregistrées à partir de 1963 jusqu'en 1980 ;
- fichier WPIL constitué par les références à partir de 1981. Ce fichier est mis à jour chaque semaine avec 6000 nouvelles inventions environ.

La base Derwent Patents (WPI/WPIL) couvre beaucoup plus de pays que la base Chemical Abstracts (mais celle-ci traite, comme la base Derwent, tous les pays importants dans le domaine de la Propriété Industrielle, c'est-à-dire déposant beaucoup de brevets).

## 2 - STRATEGIES DE RECHERCHE

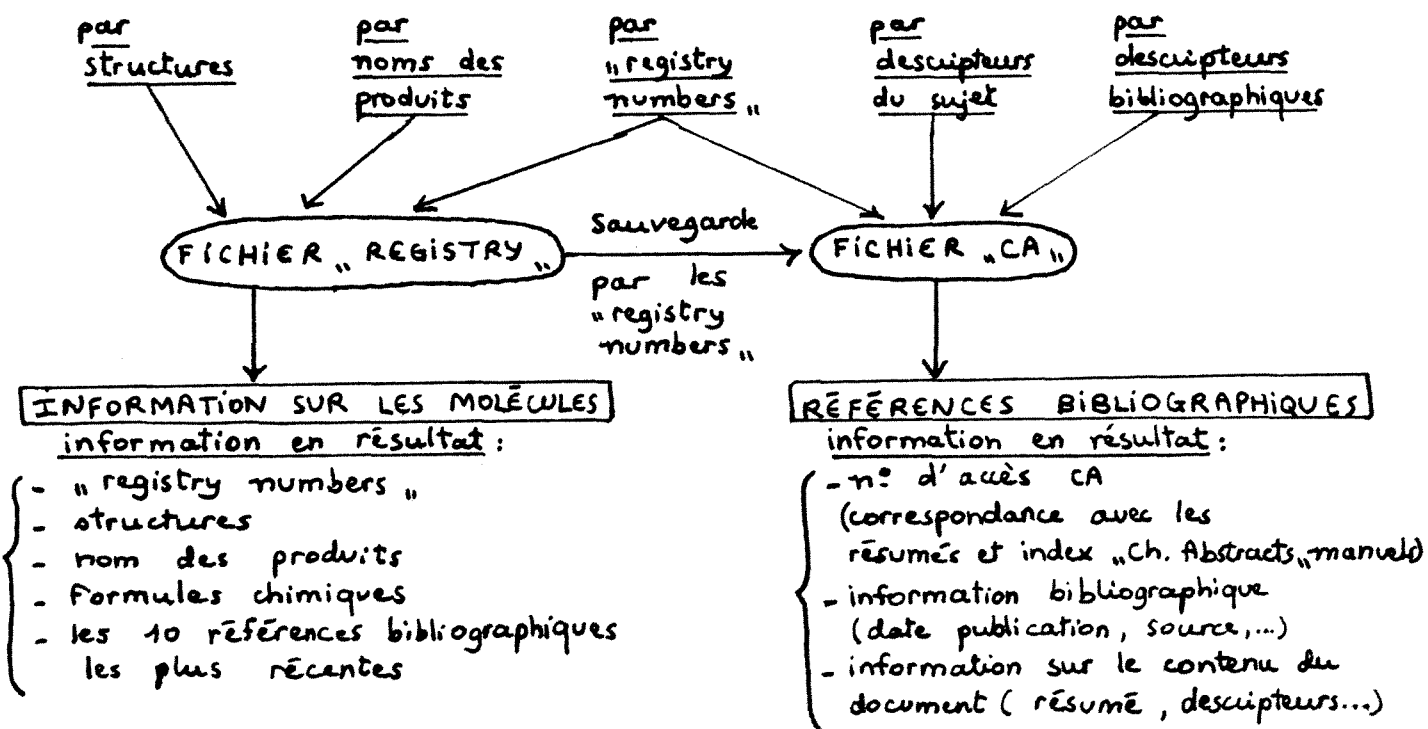
Pour une recherche de références sur un produit (une molécule) précis, il est possible d'interroger par le nom commun ou par le nom commercial, mais cette méthode est peu exhaustive. On préfère donc généralement interroger la base de données à partir de la structure de la molécule :

### 2-1 Base Chemical Abstracts

A chaque molécule est associé un numéro ("register number": RN), caractéristique et très précis (ainsi, deux isomères optiques, de propriétés différentes, ont deux numéros différents). La recherche par "register number" est très rapide (une seule étape, dans un fichier inversé) et sûre.

Il est possible également avec de nombreux terminaux, d'effectuer une recherche documentaire par construction de structures, en définissant des familles plus ou moins importantes de molécules, selon le nombre de groupes variables d'atomes à l'intérieur de la structure (système comparable au système Darc de Télésystèmes). Ce système est très performant et bien adapté aux besoins des chercheurs.

### RECHERCHE DOCUMENTAIRE SUR CAS ONLINE :



## 2-2 Base Derwent Patents

En comparaison avec la base précédente, les stratégies de recherche documentaire sur cette base offrent moins de possibilités.

Il existe depuis peu, un Registry Number caractéristique d'une molécule donnée, mais le fichier WPI n'a pas encore été mis à jour pour ce "Registry Number".

Il existe des "codes manuels" décrivant les différentes sous-structures chimiques à l'intérieur d'une molécule (cycles, hétéroatomes, liaisons particulières...) mais ce type de stratégie est très longue, fastidieuse, pouvant entraîner des erreurs : il faut combiner un grand nombre de codes pour décrire de façon complète le groupe de molécules recherchées, et de nombreux codes se recoupent, posant un problème de synonymie (car il sera nécessaire de tous les citer). La recherche documentaire par codes manuels est donc longue, elle entraîne un bruit considérable, et chose plus grave, du silence (certains brevets risquent de ne pas être retrouvés).

## 3 - REFERENCES BREVETS OBTENUES

Un exemple de référence de brevet au format maximum est donné pour chacune des deux bases, en annexe. On y trouvera également, les différents formats de visualisation des références.

Les références fournies par Derwent donnent une information plus complète sur la Propriété Industrielle :

- Outre le brevet choisi comme brevet de base (pour lequel une référence est créée), Derwent indique à la suite dans le champ PN, les numéros des principaux brevets équivalents (brevets concernant la même invention mais déposés dans d'autres pays).

Par contre la base CA n'indique comme numéro de brevet, que le numéro du brevet choisi comme brevet de base.

Ce choix d'un brevet comme brevet de base est plus ou moins arbitraire (il peut s'agir du premier brevet déposé ; ou bien si le pays de dépôt est un pays déposant peu de brevets, d'un brevet américain ou européen, par exemple...) ; les deux bases ne choisissent pas toujours, dans une famille de brevets (= ensemble des équivalents) le même brevet comme brevet de base.

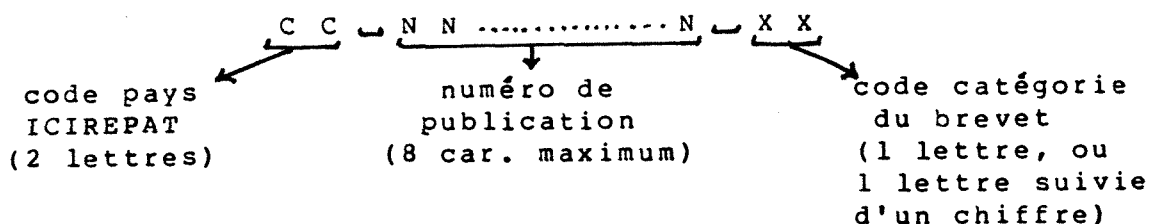
- Derwent cite généralement davantage de priorités (brevets déposés antérieurement sur la même invention d'origine).

## 4 - FORMAT DES NUMEROS DE BREVET :

Les numéros caractéristiques des brevets, ne sont pas enregistrés de la même manière dans les deux bases de données.

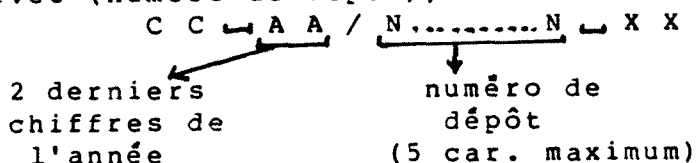
### 1 - Base "Chemical Abstracts" :

Schématiquement les numéros de brevet se présentent ainsi :



Dans le cas général, le numéro de publication est écrit sans modification, quelque soit le nombre de chiffres (pas supérieur à sept) ; le code indiquant la catégorie du brevet est facultatif.

Dans le cas particulier des pays intégrant l'année dans le numéro de brevet (numéro de dépôt), le format est le suivant :



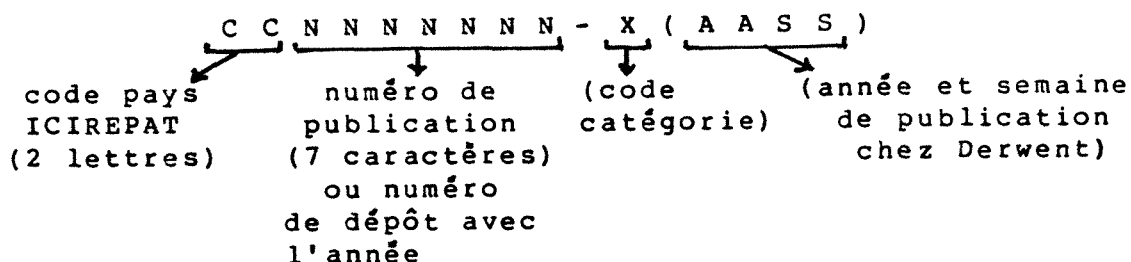
Pour le Japon, le numéro de brevet est donné sous deux formes :

- 1 - numéro de brevet avec l'année du calendrier grégorien ;
- 2 - numéro de brevet avec l'année du calendrier de l'Empereur (25 ans de moins par rapport au calendrier grégorien).

## 2 - Base "Derwent Patents" (WPI/WPIL)

Début 1984, les règles d'écriture des numéros de brevet dans la base ont été simplifiées ; en particulier, le code pays est maintenant toujours conforme (sauf pour le Japon) au code ICIREPAT. Sans ces simplifications, un programme de correspondance entre les formats d'écriture des numéros de brevet dans les deux bases n'aurait pas été réalisable.

Les numéros de brevet dans la base WPI ont un format constant de 9 caractères au total et se présentent ainsi :



Remarque : Pour une recherche en ligne par les numéros de brevet, les 9 caractères du code ICIREPAT et du numéro de publication (ou de dépôt) sont seuls utilisés.

Dans le cas général, le numéro de publication est précédé si nécessaire de tirets, de façon à avoir 7 caractères au total.

Pour les pays incluant l'année de publication ou d'application dans le numéro de brevet, les deux chiffres de l'année sont suivis si nécessaire de zéros, de façon à avoir 7 caractères pour le numéro de brevet, soit 9 caractères au total.

- Japon : le code ICIREPAT (JP) est modifié en J ; l'année de publication est soit l'année du calendrier grégorien (brevets "koho", publiés après accord), soit l'année de l'Empereur (brevets "kokai", publiés après 18 mois).
- Hongrie : le code ICIREPAT (HU) est suivi de la lettre T (brevets

publiés après examen et accord) ou de la lettre H (autres catégories de brevets), des zéros précédant au besoin le numéro de publication, de façon à avoir 9 caractères au total.

## VI/ CHOIX D'UNE APPLICATION

### 1 - INTERROGATIONS. EMISSION ET RECEPTION DES DONNEES

Pour pouvoir émettre des données (numéros de brevets) vers la base de données Derwent, et mémoriser les références provenant de chacune des bases interrogées, les interrogations sont faites non pas par l'un des terminaux simples (sans mémoire) habituellement utilisés, mais au moyen du miniordinateur WANG OIS, en utilisant le terminal de traitement de texte équipé d'un Modem, qui se trouve au Service Documentation.

L'émission ou la réception de données est faite par lancement d'une session interactive de transmission (liaison TTY, vitesse 300 bauds), à partir du terminal.

Si la réception de données (références de brevets) dans un "Document" (sur le Disque-système) créé automatiquement, ne posait pas de problèmes, il n'en allait pas de même pour l'émission d'un Document (contenant les numéros de brevet combinés par l'opérateur booléen "ou" en une étape de recherche) au cours de l'interrogation de la base Derwent.

En effet les procédures de transmission et l'envoi de l'information contenue dans le Document prennent un certain temps ; cependant les essais menés ont été concluants, la base Derwent reste connectée, attendant la question de l'utilisateur jusqu'à ce que la totalité du Document soit reçue.

### 2 - FICHER RESULTAT

Les références brevets obtenues par interrogation de chaque base de données, sont mémorisées dans deux Documents traitement de texte, sur le Disque-système. Il est intéressant, pour conserver ces références, de les réunir dans un fichier bibliographique interne, avec une référence brevet par enregistrement, afin de pouvoir réaliser différents types d'édition, et de recherche documentaire sur ces références. Entre la création d'un fichier Basic, et la création d'un fichier Texto, le choix s'est porté sur le fichier Texto.

En effet ce logiciel documentaire, déjà utilisé pour les fichiers internes du Service Documentation, est très complet et offre de larges possibilités :

- fichiers à champs de longueur variable (nécessité pour les références brevets, aux champs parfois très longs) ; possibilités d'index divers, de tris, de constitution de sous-fichiers ; facilités de mise à jour (en particulier par "ajout piloté" automatique des références contenues dans un Document traitement de texte) ; multiples possibilités de sélections de références et de formats d'édition, éventuellement programmées par "Logotext"...

Les possibilités de programmer (Logotext) des "dialogues" d'édition et d'interrogation, ont semblé particulièrement intéressantes, permettant éventuellement une étude des références brevets enregistrées, pouvant servir à établir des stratégies de recherche en chimie, ou à comparer les résultats des deux bases de données.

### 3 - PROGRAMMATION

La programmation a été nécessaire à deux niveaux de l'application:

- Pour sélectionner les numéros de brevet dans les références Chemical Abstracts mémorisées, les réécrire sous la forme Derwent, les combiner par l'opérateur booléen "ou" en une étape de recherche de façon à constituer le Document à émettre lors de l'interrogation de la base Derwent ;
- Pour traiter les références brevets Chemical Abstracts et Derwent obtenues en résultat de l'interrogation des deux bases de données, de manière à ne conserver que les champs intéressants, fusionner les références CA et Derwent traitant de la même famille de brevets (même numéro de brevet de base ou équivalent), et présenter les références en vue de la mise à jour automatique (ajout piloté) du fichier Texto.

Entre les deux possibilités offertes par le système Wang OIS : programmation en traitement de texte ou programmation en Basic, on a choisi les lexiques (= programmes) de traitement de texte.

En effet, le Basic Wang, quoique orienté vers le traitement de texte, présente l'inconvénient de procédures assez complexes pour la création et la manipulation de fichiers, et pour la conversion de Documents en fichiers.

La création de "lexiques" de traitement de texte, a semblé présenter des possibilités de programmation suffisantes pour traiter le problème considéré.

Un lexique présente l'inconvénient d'être beaucoup plus long à l'exécution qu'un programme Basic, mais l'accès au Document à traiter y est très simple, et les fonctions sont bien adaptées à la réécriture des numéros de brevet (insertion ou suppression de caractères essentiellement), ainsi qu'à la présentation de références pour "ajout piloté" sur fichier Texto.

### 4 - ENCHAINEMENT DES TRAVAUX

L'interrogation des deux bases de données est faite au Service Documentation, au moyen du terminal Wang OIS relié à un Modem.

Lors de chaque interrogation, les références sont mémorisées dans un Document créé automatiquement sur le Disque-système (document "A" pour les références Chemical Abstracts, et document "B" pour les références Derwent).

Trois lexiques sont utilisés pour traiter les références de brevets, les lexiques n°1 (0283), n°2 (0297) et n°3 (0294).

L'exécution de ces lexiques fait intervenir, outre les documents "A" et "B" contenant les références, trois documents déjà créés (1041D, 1042D, 1043D), vides au départ, et vides à nouveau après l'exécution des lexiques (en vue de leur réutilisation à la prochaine recherche).

On obtient en résultat :

1°/ Dans le document "A" :

- la liste des références CA d'origine résultant de la première interrogation, en page 1 ;
- la liste des références Derwent d'origine obtenues lors de la deuxième interrogation, en page 2 ;
- l'étape de recherche (numéros de brevet combinés) telle

qu'elle a été émise vers la base Derwent, en page 3.

Ce document (auquel on donne comme nom le groupe de molécules étudié) est archivé sur une disquette. On peut ainsi sauvegarder les références d'une ou plusieurs recherches, temporairement, par sécurité (en cas d'erreur commise sur les numéros de document lors de l'exécution des lexiques, par exemple).

2°/ Dans le document "B" :

Les références Chemical Abstracts et Derwent traitées, présentées en vue de l'ajout piloté sur le fichier Texto.

Ce document est détruit, une fois la mise à jour du fichier Texto réalisée.

Les étapes successives à effectuer sont les suivantes :

### 1°/ Interrogation de la base Chemical Abstracts :

L'utilisateur lance une session interactive à partir du terminal Wang OIS. Puis il se connecte au réseau Transpac et au serveur Cas Online. La recherche documentaire sur le groupe de molécules étudié est effectuée. Soit Ln la dernière étape de recherche sur le fichier bibliographique CA ; il crée une nouvelle étape : Ln AND P/DT (recherche sur champ "type de document" = brevet) de manière à sélectionner les brevets.

Puis il se met en mode "ouverture réception" et demande dans une nouvelle étape la visualisation des références en format "BIB". Une fois cette visualisation terminée, il se met en mode "fermeture réception" (le document "A" créé automatiquement au moment de l'ouverture en mode réception, et contenant ces références, est alors refermé). Il ne reste plus alors qu'à se déconnecter du serveur CAS Online, puis de Transpac, puis à stopper la session interactive en revenant au menu "traitement de texte" habituel.

### 2°/ Exécution du lexique n° 1 sur les références CA

Le lexique n° 1 (0283) est exécuté sur le document "A". Le texte inutile est ainsi effacé, seules les références sont conservées.

On obtient, dans le document 1043D, la liste des numéros de brevet tirés des références CA, combinés par l'opérateur booléen "ou" (représenté par un point) en une étape de recherche (selon le logiciel Orbit de SDC). Par ailleurs, le document 1041D contient les références partiellement modifiées (en vue de l'exécution du lexique n° 3), et le document "A" les références CA d'origine, telles qu'elles ont été mémorisées.

### 3°/ Interrogation de la base Derwent Patents

L'utilisateur lance comme précédemment, une session interactive, puis se connecte sur le réseau Transpac et le Serveur SDC.

Il demande le fichier WPI (FILE WPI), puis déclare le point synonyme de l'opérateur booléen "ou" (SYNONYM . FOR OR).

Puis il se met en ouverture du mode "émission", et donne le numéro du document à émettre (1043D). Les numéros de brevet sont ainsi automatiquement demandés, sans avoir à les taper au clavier.

On reçoit en réponse le nombre de références ; comme précédemment l'utilisateur se met alors en mode réception (un document "B" est créé automatiquement) et demande la visualisation des références en format standard (PRINT).



Puis la recherche est sauvegardée, et rappelée sur le fichier WPIL. Les références obtenues sur ce fichier sont visualisées ; le document est alors refermé (fermeture du mode réception) et l'utilisateur déconnecte le terminal de SDC et de Transpac, puis arrête la session interactive.

4°/ Exécution du lexique n° 1 sur les références Derwent :

Le même lexique (0283) est appliqué aux références Derwent contenues dans le document "B".

On obtient, dans le document 1042D, les références partiellement modifiées, et dans le document "B", les références d'origine (le texte inutile ayant été supprimé).

5°/ Exécution du lexique n° 2 :

Ce lexique réunit dans le document "A", les références Chemical Abstracts et Derwent d'origine, ainsi que l'étape de recherche (contenu du document 1043D).

6°/ Archivage :

Le document "A" est alors sauvegardé sur disquette.

7°/ Exécution du lexique n° 3 :

Ce lexique s'exécute sur les documents 1041D et 1042D ; on obtient en résultat dans le document "B" (vide à la suite de l'exécution du lexique n° 2), les références pour l'ajout piloté sur le fichier Texto.

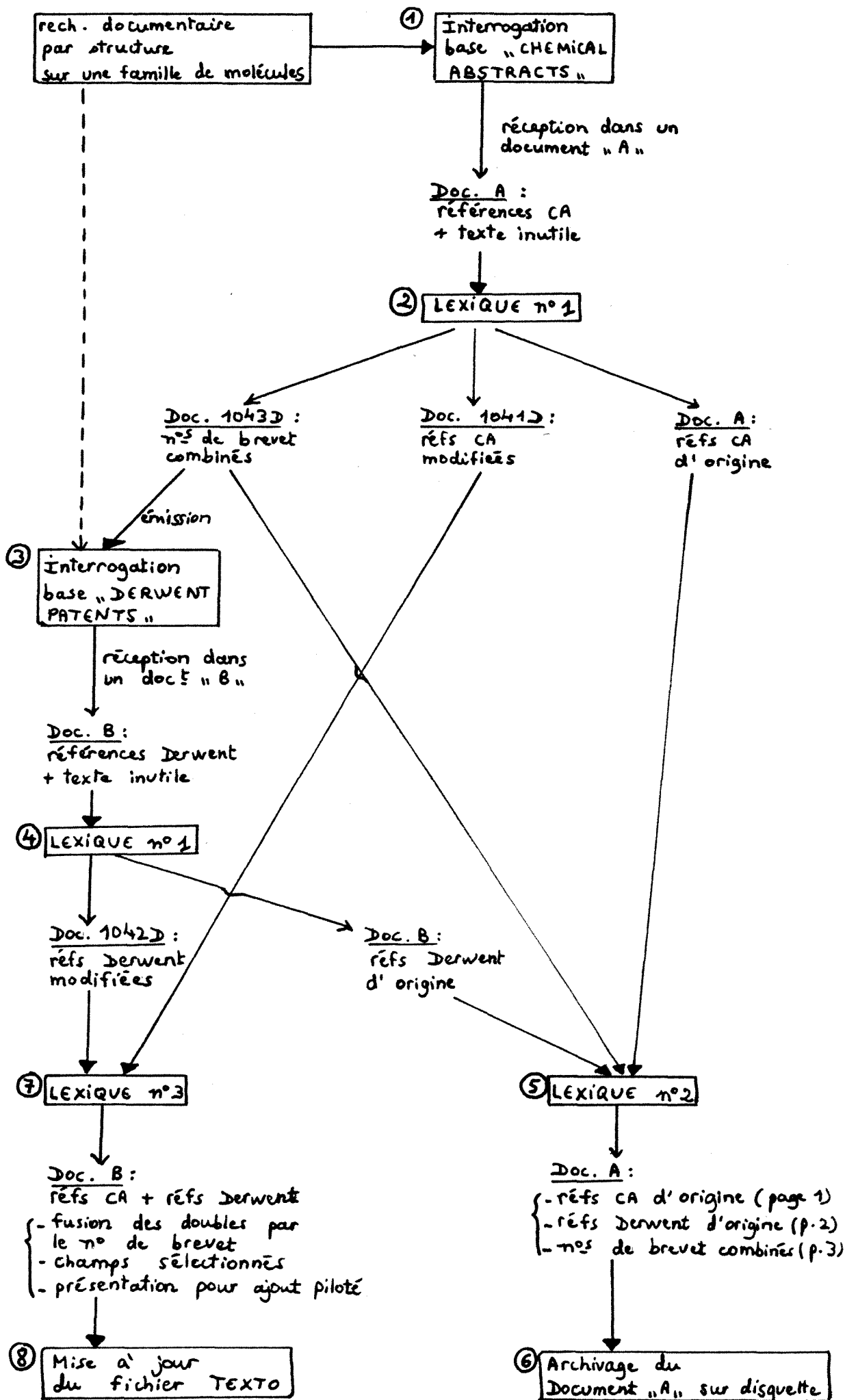
8°/ Mise à jour du fichier Texto :

Celle-ci est effectuée très simplement par ajout piloté à partir du document "B".

Les index éventuels sont mis à jour automatiquement.

Ces différentes étapes, en apparence longues et complexes mais assez simples dans la pratique, sont résumées sur le schéma page suivante.

# ENCHAÎNEMENT DES TRAVAUX



PROCÉDURES À SUIVRE POUR L'EXÉCUTION  
DES LEXIQUES 1, 2, 3 sur le terminal de traitement de texte WANG a.i.s:

AFFICHÉ A L'ECRAN :

A TAPER PAR L'UTILISATEUR :

menu "traitement de texte"	- Sélection Fonction Lexique , puis <u>EXECUTEZ</u>	Lexique n°1
(rattachement du <u>lexique 1</u> )	- <u>EXECUTEZ</u> puis: 0283␣	
(lancement de l'exécution)	- <u>GL</u> A	
"n°Doc Référence ?"	- n°du Doc A (du Doc B) puis: <u>EXECUTEZ</u>	
(lexique en cours)		
"n°Doc Résultat ?"	- 1041D ( 1042D ) puis: <u>EXECUTEZ</u>	
"n°Doc Référence ?"	- n°du Doc A (du Doc B) puis: <u>EXECUTEZ</u>	
(lexique en cours)		
"lexique terminé"	- <u>ANNULATION</u> puis <u>EXECUTEZ</u>	

menu "traitement de texte"	- Sélection Fonction Lexique, puis <u>EXECUTEZ</u>	Lexique n°2
(rattachement du <u>lexique 2</u> )	- <u>EXECUTEZ</u> puis: 0297␣	
(lancement de l'exécution)	- <u>GL</u> A	
"n° Doc A ?"	- n° du Doc A, puis: <u>EXECUTEZ</u>	
"n° Doc B ?"	- n° du Doc B, puis: <u>EXECUTEZ</u>	
(lexique en cours)		
"lexique terminé"	- <u>ANNULATION</u> puis: <u>EXECUTEZ</u>	

menu "traitement de texte"	- Sélection Fonction Lexique, puis <u>EXECUTEZ</u>	Lexique n°3
(rattachement du <u>lexique 3</u> )	- <u>EXECUTEZ</u> puis: 0294␣	
(lancement de l'exécution)	- <u>GL</u> A	
"n° Doc B ?"	- n° du Doc B, puis: <u>EXECUTEZ</u>	
(lexique en cours)		
"lexique terminé"	- <u>ANNULATION</u> , puis <u>EXECUTEZ</u>	

Remarque : Pour le lexique n° 1 :

- s'il est exécuté après l'interrogation de la base Chemical Abstracts:  
tapez le n° du Doc A (= Document créé automatiquement pendant l'interrogation  
et contenant les références CA) comme Document Référence ;  
tapez 1041D comme Document résultat.

- s'il est exécuté après l'interrogation de la base Derwent Patents:  
tapez le n° du Doc B (= Document créé automatiquement pendant l'interrogation  
et contenant les références Derwent) comme Document Référence ;  
tapez 1042D comme Document résultat.

## VII/ PROGRAMMES DE TRAITEMENT DE TEXTE

La fonction "lexique" sur le système WANG OIS, permet de programmer en traitement de texte, c'est-à-dire de chaîner les fonctions (supprimer un caractère, rechercher une séquence de caractères, recopier un document,...) utilisées habituellement en traitement de texte manuel au clavier.

Les principales fonctions "lexique" sont données en annexe.

On constitue ainsi une succession d'entrées (sous-programmes) qui s'enchaînent automatiquement.

Il est possible de réaliser des boucles et des bifurcations en fonction d'une condition réalisée ou non (test (-SI-)) ; cependant les possibilités de programmation sont tout de même inférieures à celles des langages évolués tels que le BASIC.

De nombreux lexiques sont utilisés au Service Documentation, pour la gestion des fichiers internes (présentation de documents pour ajout piloté sur Texto,...).

Les inconvénients majeurs rencontrés avec la programmation de lexiques (programmes) en traitement de texte Wang OIS ont été les suivants :

- pas de possibilité de numéroter les références pour les distinguer ; il est possible de noter un emplacement par une "note", mais le marquage d'une note efface la précédente ; de plus c'est l'emplacement dans le document qui est noté et non un caractère particulier (si une ligne est supprimée la note est déplacée) ; ceci fait que le document est très souvent relu dans son intégralité, ce qui allonge le temps d'exécution et n'est pas intellectuellement satisfaisant.

- Il n'est pas possible de rechercher un numéro de brevet (= une séquence de caractères), puis un autre,... successivement, à moins d'utiliser le subterfuge suivant : copie du numéro à sa droite, puis remplacement global de ce numéro dans le document par un symbole, à partir de sa copie ; enfin recherche de ce symbole dans le document. Or en Basic il suffit de copier les numéros dans un vecteur, pour utiliser successivement ces numéros dans l'exécution du programme.

- Il faut fréquemment effectuer des tests "bas d'écran", "haut d'écran", "fin de document". Ainsi un "espace-arrière" ne sera pas pris en considération si le curseur est en haut d'écran. Ceci allonge et complique l'exécution du lexique, par exemple pour remplacer des séquences comportant un retour à la ligne.

- Le temps d'exécution est long en comparaison avec un programme Basic par exemple, et sera multiplié très rapidement avec le nombre de références, du fait du grand nombre de recherches séquentielles effectuées dans le document.

Cependant l'accès aux documents à traiter est aisé, les fonctions de traitement de texte sont bien adaptées au traitement des références (suppression de champs, insertion de retours à la ligne,...) et le temps d'exécution est acceptable pour un nombre de références compris entre 50 et 200 environ.

Les trois lexiques créés sont donnés en annexe, avec les

organigrammes correspondants.

D'une façon générale pour les lexiques n° 1 et 2, lorsqu' un numéro de document est demandé à l'utilisateur, il l'est par la fonction (-N-TOUCHES-) qui permet une validation : le document correspondant au numéro frappé apparaît, ce qui permet une vérification avant de relancer la poursuite du lexique en frappant la touche (-EXECUTEZ-).

## 1 - LEXIQUE N° 1

Ce lexique s'applique soit sur le document "A" contenant les références CA, soit sur le document "B" contenant les références Derwent. (Il pourrait également s'appliquer à un document contenant des références des deux bases en désordre).

Présentation de l'étape de recherche (numéros de brevet combinés) :

La base Derwent Patents est interrogée par l'intermédiaire du serveur SDC, au moyen du logiciel ORBIT.

Pour une recherche par les numéros de brevets, le nom du champ inversé en lexique (numéro de brevet : /PN )est implicite.

Une longue instruction de recherche, avec des termes (les numéros de brevet) combinés par des opérateurs (l'opérateur booléen "ou") peut être écrite sur plusieurs lignes, en terminant chaque ligne par un opérateur ; l'étape ne doit pas comporter plus de 380 termes ou opérateurs, ni plus de 5800 caractères.

Le logiciel ORBIT permet d'établir, au début d'une recherche, un synonyme pour une commande ou un opérateur donné. On déclare, au début de l'interrogation, le point synonyme du "ou".

On peut donc combiner, en une seule étape, 190 numéros de brevet maximum, sur 32 Lignes (9 caractères par numéro ; 3 caractères par opérateur : espace-point -espace ; une ligne faisant 80 caractères, on a donc 6 numéros par ligne). Cette limite est apparue suffisamment large. Pour l'envoi d'un nombre supérieur de numéros de brevet, il faudrait scinder l'étape de recherche dans deux documents (soit 380 numéros maximum) ou plus.

Après avoir demandé le numéro du document à traiter, le lexique modifie le format du document en vue de la réécriture des numéros de brevet tels qu'ils sont enregistrés dans la base Derwent ( pour les entrées Q à V) (entrée A) ;

Les références de l'une ou l'autre base se distinguent par le fait qu'elles commencent par la séquence (-RETOUR-)AN , et s'achèvent par une ligne blanche, soit un symbole (-RETOUR-) à la marge gauche. Le reste du texte est supprimé, et les références sont séparées par la séquence //(-ALINEA-) en vue de l'ajout piloté final (B, C, D).

L'entrée E referme le document, et l'entrée F demande le numéro du document résultat, où les références (G) sont recopiées (si le document résultat n'est pas vide, il y a erreur et le lexique s'arrête, il faut alors recommencer l'entrée F).

Les champs constamment retrouvés dans les références Derwent au format standard et sélectionnés (numéro d'accès, titre, déposant numéros de brevet) sont marqués par un alinéa pour les distinguer des champs à supprimer. En ce qui concerne le champ PN, il est remplacé par une virgule ; en effet dans le lexique n° 3 la recherche d'un numéro de brevet est en fait la recherche d'un numéro

précédé d'une virgule, afin de distinguer les numéros recherchés (champ PN), des numéros du champ PD (date de publication). Le point-virgule séparant les numéros de brevet équivalents est remplacé par une virgule (séparateur d'articles habituellement utilisé sur Texto) en tenant compte des retours à la ligne (entrées H, I, J).

L'entrée K teste s'il y a des références CA. Si oui, une page est créée pour copier les numéros de ces références.

En L, le numéro de brevet de la référence CA est déplacé à côté du numéro d'accès.

Pour chaque référence, le numéro est copié en page 2. Pour les pays pour lesquels Derwent prend comme numéro le numéro de dépôt, le champ priorité est copié également, puis seuls sont conservés le code du pays et le numéro de dépôt (M, N, O). En ce qui concerne le Japon, pour lequel la base CA donne les deux numéros de brevet, quoiqu'un seul corresponde à la référence, les deux numéros sont copiés et traités en page 2 (Q, R).

Selon le cas et si besoin, des zéros ou des tirets sont insérés de manière à écrire le numéro comme dans la base Derwent. Les numéros Derwent ayant un format constant de 9 caractères au total, on teste si cette longueur est atteinte en vérifiant si le curseur est ou non sous un caractère (-TAB-) du format.(entrées Q, R, S, T, U, V). Le numéro est alors recopié dans la référence, retrouvée directement grâce à une marque (note).

Les numéros en page 2 sont alors reliés par la séquence . , sauf pour le dernier numéro (X). Cette page est alors déplacée dans le document 1043D, prêt pour l'émission vers la base Derwent (Y, Z).

L'étape de recherche par les numéros de brevet est ainsi construite automatiquement, à partir des références "Chemical Abstracts".

## 2 - LEXIQUE N° 2

Ce programme très simple réunit les références d'origine et l'étape de recherche, en vue de l'archivage. Il n'était pas indispensable mais peut limiter les erreurs manuelles toujours possibles.

## 3 - LEXIQUE N° 3

Ce lexique recopie les documents intermédiaires 1041D et 1042D (références CA et Derwent partiellement traitées) dans le document "B" (vidé par le lexique précédent).

Les références CA sont suivies des références Derwent.

Pour chacune des références CA prises successivement, le numéro de brevet converti (champ &&) est recherché dans les références Derwent.

Dans ces références Derwent, le numéro de brevet de base se trouve aussi dans le champ PD (date de publication). Pour distinguer les numéros de brevet du champ PN uniquement, on fait donc une recherche du numéro précédé d'une virgule (cf. le lexique n°1). Ceci complique le programme, car il faudra tenir compte des numéros en début de ligne (la séquence à rechercher est alors: , retour espace espace espace espace espace espace numéro ).

Pour rechercher automatiquement cette séquence, la fonction recherche totale n'était pas applicable. On recopie donc cette séquence (,n°), puis on effectue un remplacement global de cette copie dans la suite du document, par le symbole §

Le premier symbole du document (qui remplace la copie du numéro) est supprimé, puis ce symbole est recherché dans les références Derwent. S'il n'est pas retrouvé, on recommence le même processus avec la séquence ,~~...n°~~ pour le cas où le numéro cherché se trouverait en début de ligne. Si le symbole n'est pas trouvé à cette deuxième recherche, c'est donc que le numéro de brevet de la référence CA n'a pas été retrouvé dans les numéros d'équivalents des références Derwent : la référence CA est donc conservée.

Lorsque le symbole § est retrouvé dans une référence Derwent, sa position est notée. Puis le numéro d'accès CA et le numéro de brevet converti sont déplacés de la référence CA vers la note. Le symbole est ainsi remplacé par le numéro d'origine, et le numéro d'accès CA est placé au-dessus du champ PN. La référence CA est alors supprimée, et le lexique traite la référence suivante.

Lorsque toutes les références CA ont été traitées, (donc toutes celles "en double" supprimées, en fusionnant le numéro d'accès CA avec la référence Derwent correspondante), les références sont alors traitées pour l'ajout piloté (remarque : ce lexique pourrait traiter un document ne contenant que des références CA, ou que de références Derwent) :

Les champs intéressants sont marqués par un alinéa.

Les entrées S et T marquent des alinéas sur les lignes des champs comportant plusieurs lignes.

Les lignes ne comportant pas d'alinéa sont alors supprimées (la première ligne du document : //→◀ quoique comportant un alinéa, est automatiquement supprimée, car elle ne doit pas exister pour l'ajout piloté).

Dans le cas d'un champ à plusieurs lignes, un tiret est placé en fin de ligne (remplacement des séquences retour alinéa par des séquences tiret retour : entrée V) pour indiquer à Texto que le champ n'est pas terminé.

Les séquences //→◀ sont remplacées globalement par //◀

Ce double slash indique à Texto la fin d'une référence.

Enfin l'alinéa est remplacé globalement dans tout le document par un retour à la ligne.

#### Remarques :

##### 1°/ Présentation pour l'ajout piloté :

On obtient finalement la présentation suivante, pour la mise à jour automatique du fichier Texto :

```
nom de champ
contenu du champ      1ère référence
nom de champ
contenu du champ
...
//
nom de champ          2ème référence
contenu du champ
...
//
...etc
```

L'ordre des champs n' a pas d'importance.

Il n'est pas obligatoire de mentionner tous les champs d'une référence (le champ sera alors enregistré comme vide).

Le champ indiquant le numéro de référence sur le fichier sera créé et rempli automatiquement, dans l'ordre de la saisie.

## 2°/ Suppression des références CA "en double" :

Il peut arriver que plusieurs références CA traitent de la même famille de brevet, mais ceci semble assz rare dans la pratique.

Sur la base Derwent une seule référence est créée par famille de brevets.

Il peut arriver qu'un brevet choisi comme brevet de base par Chemical Abstracts, ne soit pas inclus dans les équivalents choisis par Derwent, pour une référence traitant cependant de la même famille de brevets. Dans ce cas on aura donc deux références sur le fichier Texto. Cependant un tel cas est assez peu probable dans la pratique, car le brevet de base choisi par la base CA sera généralement un brevet important, donc pris également en compte chez Derwent au niveau des équivalents (Derwent cite en général un grand nombre d'équivalents).

On a étudié s'il était possible de supprimer les doubles une fois le fichier Texto mis à jour, soit par questions combinées, soit par création d'index, soit par création de sous-fichiers. On aurait alors pu enchaîner les opérations à effectuer dans un ou plusieurs dialogues (programmes) successifs, et la suppression des références en double d'après le numéro de brevet de base ou équivalent aurait été ainsi effectuée très simplement et rapidement.

Cependant il a semblé que ce n'était pas réalisable.

Si l'on crée un index sur le numéro de brevet, Texto indiquera éventuellement les numéros des références en double, mais il faudra alors les supprimer manuellement, ce qui n'est pas concevable, vu que par définition le nombre de doubles sera élevé.

Il peut être intéressant, lorsque la base Derwent est interrogée, d'effectuer, en plus de l'émission des numéros de brevet CA, une recherche documentaire par structure, pour essayer d'obtenir des brevets supplémentaires (propres à la base Derwent), la base Derwent étant plus complète et couvrant beaucoup plus de pays que la base CA. Puis ces deux étapes de recherche sont combinées, de manière à n'obtenir qu'une série de références en résultat de l'interrogation (l'intérêt de passer de la base CA à la base Derwent par les numéros de brevet étant d'obtenir des références plus complètes avec l'accès au texte intégral des brevets, par le numéro d'accès Derwent).



## VIII/ FICHER RESULTAT

Le fichier Texto contenant les références traitées n'a pas été constitué faute de temps.

On peut cependant définir ses caractéristiques :

On choisit une numérotation automatique des références, avec un format de numérotation de 4 caractères et un pas de 1.

A partir des références Chemical Abstracts et Derwent, ont été sélectionnés les champs suivants, définis implicitement comme de longueur variable :

(ref) : numéro de la référence (attribué automatiquement par le logiciel)  
CH : numéro d'accès CA  
WP : numéro d'accès Derwent  
TI : titre  
PA : nom du déposant  
PI : numéro de brevet CA  
PN : numéro(s) de brevet Derwent  
CT : numéro(s) des brevets cités  
PD : date de publication  
PR : date(s) de priorité et numéro(s) de brevet correspondant

Ces champs peuvent être vides. Dans le cas d'une fusion entre une référence CA et une référence Derwent, Les seuls champs CA conservés sont le numéro d'accès et le numéro de brevet ; pour les champs communs aux deux bases de données, le contenu des champs Derwent est choisi, étant plus complet.

Le champ PN est éclaté en articles (les numéros de brevet), le séparateur étant la virgule.

On constitue un index sur le champ PN (en tronquant les termes à 9 caractères, de façon à avoir un index sur le numéro de brevet), index éclaté par articles, et mis à jour automatiquement en même temps que le fichier des références.

En questionnant sur ce champ PN avec troncature après les deux premières lettres (code ICIREPAT, on pourra sélectionner les brevets selon le pays).

Il serait aussi possible de demander les références contenant le champ CH non vide mais ayant le champ WP vide, afin de voir la fréquence des références Chemical Abstracts non retrouvées sur la base Derwent.

Plusieurs recherches peuvent être enregistrées à la suite sur le même fichier ; pour les différencier, il suffit de créer un champ supplémentaire, contenant le numéro d'ordre de cette recherche et rempli automatiquement dans toutes les références d'une recherche par commande combinée.

## IX/ CONCLUSION

L'application qui a été réalisée (partiellement puisque les fichiers et index Texto n'ont pas été créés faute de temps) est conçue pour un nombre de références jusqu'à 150-200 environ, et constitue plutôt un stade expérimental.

Au-delà, il faudrait envisager un autre type d'application (fichiers Basic avec index sur numéro de brevet,...)).

Les inconvénients d'un tel système sont essentiellement la nécessité de visualiser toutes les références en ligne (ce qui est long et coûteux) et la lenteur du traitement des références par les lexiques de traitement de texte.

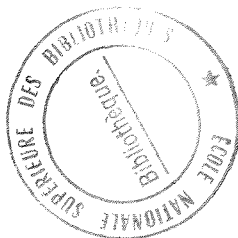
Les avantages en sont la conservation, l'organisation et le traitement de l'information contenue dans les références de brevet, de manière à en dégager l'essentiel.

De telles applications constituant des fichiers de références en vue d'un traitement et d'une étude de l'information contenue dans ces références, sont sans doute appelées à se développer.

Dans le cadre de ce stage j'ai pu observer l'application réalisée au Centre de Recherche Rhône-Poulenc St-Fons : de très nombreuses références brevet sont ainsi enregistrées dans divers fichiers et étudiées de manière approfondies par le logiciel de statistiques SAS pour mettre en évidence les caractéristiques des références et les comparer sur un grand nombre.

Ce type de documentation véritablement stratégique qui consiste donc à analyser d'une manière approfondie les références obtenues par l'interrogation de bases de données externes, peut être particulièrement intéressant pour définir des orientations au niveau de la recherche en chimie, et améliorer la recherche documentaire elle-même.

En conclusion, je voudrais dire que ce stage m'a intéressée par sa diversité : il m'a permis en effet de travailler sur divers aspects de la documentation : Propriété Industrielle, Interrogation de bases de données externes, Traitement de texte, Fichiers de documentation internes, et de voir le fonctionnement d'un Service Documentation important, utilisant les possibilités de l'Informatique documentaire.





\*9518059\*