



## ***Système de Gestion Electronique de Documents***

*Version 1.0*

---

# intellia

3, Rue Mohamed Abdou 20000 Casablanca - Maroc  
Tél. : 212 (0) 22 29 37 70 / 22 29 43 30 - Fax : 212 (0) 22 29 60 95  
79, Avenue Ibn Sina, Agdal 10000 Rabat - Maroc  
Tél. : 212 (0) 37 67 21 91 / 67 21 92 - Fax : 212 (0) 37 67 21 93  
e-mail : [info@intellia.ma](mailto:info@intellia.ma)

|   |    |
|---|----|
| ▪ Avant Propos .....                        | 1  |
| ▪ Le Concept OneDoc.....                    | 3  |
| ▪ L'Environnement OneDoc.....               | 4  |
| ▪ L'Architecture de OneDoc.....             | 6  |
| ▪ Description Fonctionnelle de OneDoc ..... | 8  |
| ▪ Spécifications Techniques de OneDoc.....  | 22 |

Ce document constitue une présentation des spécifications techniques et fonctionnelles du **Système d'Information Documentaire** (Document Management System) **OneDoc**, développé par **OneCor Software**.

Le logiciel OneDoc est architecturé autour d'une entité maîtresse : le **document**. Dans OneDoc, la notion de document est prise au sens large du terme et désigne tout objet de connaissance ou support d'information quelle que soit sa nature :

- Support physique (tout document imprimé ou support d'information tangible)
- Support virtuel (fichiers ou ensemble de fichiers)
- Référence à un document externe.

L'objectif du logiciel OneDoc est de constituer, d'enrichir, de gérer et de maintenir le patrimoine de connaissances de l'entreprise.

La modularité et la richesse fonctionnelle du logiciel OneDoc lui permettent de s'adapter à tous les besoins de l'entreprise :

- Gestion documentaire : gestion d'un centre de documentation, bibliothéconomie, etc.
- Gestion électronique des documents (GED) ou ensemble d'outils et de techniques assurant la dématérialisation, le classement, le stockage et l'accès aux documents :
  - GED Administrative (Dossiers clients, fournisseurs, ...).
  - GED Bureautique (Courriers, notes internes, circulaires, ...).
  - GED Financière (Factures, BL, bons de commande, ...).
  - GED Technique (Plans, notes techniques, ...).
- Gestion de banques d'images fixes : Photothèques.
- Applications de e-business (Internet, Intranet, Extranet, Web et Wap)
  - Moteur de recherche et Annuaire de pages Web.
  - Journal en ligne.
  - Site de commerce électronique doté d'un catalogue en ligne.

Les entités ou objets documentaires et les supports d'informations gérés par OneDoc sont de deux types :

## Entités Documentaires

- **L'entité documentaire primaire de type document**, qui se compose de données de forme et de nature très variables :
  - Forme : Texte, tableau, image, graphique, dessin, son, image vidéo, multimédia
  - Nature ou Type : notice documentaire, fichier, image numérisée, fax, donnée Internet, message électronique, dépêche d'agence, donnée personnalisée obtenue par technologie de diffusion dynamique d'informations actualisées (« *push* »)
- **L'entité documentaire virtuelle** de type dossier, qui peut contenir d'autres dossiers ou des documents.

## Sources

Les entités documentaires gérées par OneDoc peuvent provenir de sources multiples :

- Communication Télex,
- Réception satellite,
- Passerelle fax,
- Passerelle messagerie,
- Volume de disque dur,
- Scanner,
- Caméra,
- Internet,
- Téléphone, etc.

## Acquisition

Les documents peuvent être intégrés dans le système de façon automatique, semi-automatique ou manuelle.

Les fonctions de traitement en amont (*pre-processing*) réalisées par OneDoc sont les suivantes :

- Acquisition par capture papier ou capture électronique (numérisation, réception, lecture, etc.)
- Analyse
- Indexation
- Distribution
- Stockage dans la base de données et/ou sur disque.

Les fonctions de traitement en aval (*post-processing*) sont les suivantes :

- Archivage
- Modification
- Suppression

L'entité documentaire est constituée de trois composantes fondamentales:

- Les Propriétés : Titre, date, auteurs, mots clés, etc.
- La Structure : liens hiérarchiques, liens de rattachement, etc.
- Le Contenu : localisation dans la base de donnée, dans un fichier externe, dans un document physique ou sur Internet.

Dans sa version de base, OneDoc est constitué d'une partie Serveur, **OneDoc Server**, et d'une partie Cliente, **OneDoc Client**.

### OneDoc Server

La partie serveur du logiciel OneDoc se compose essentiellement du serveur de base de données relationnelles et du moteur d'indexation et de recherche en texte intégral. **OneDoc Server** contient les modules fonctionnels qui constituent le cœur du Système de gestion documentaire et de gestion électronique de documents en architecture Web Serveur :

- **OneDoc Document Server :** Serveur documentaire
- **OneDoc WebDAV Server :** Serveur de fichiers
- **OneDoc Image Server :** Serveur d'images
- **OneDoc Search Server :** Serveur d'indexation et de recherche en texte intégral.

### OneDoc Client

La partie cliente du logiciel OneDoc, **OneDoc Client**, est conçue en architecture totalement Web. Elle se compose de cinq modules de base :

- **OneDoc Explorer**  
C'est le module de consultation et d'exploitation de la base de données documentaire. Il permet l'utilisation des différentes applications de production de l'information documentaire : Saisie des notices, Thésaurus, Logiciels Office, etc.
- **OneDoc Imaging**  
Ce module permet la numérisation, le traitement et l'archivage de documents images.
- **OneDoc Manager**  
C'est le module d'administration de l'environnement OneDoc.
- **OneDoc Designer**  
Ce module permet à l'administrateur du système de réaliser la création et la modification des écrans OneDoc, ainsi que l'ajout et la modification des tables ou champs de la base de données documentaire.
- **OneDoc Builder**  
Ce module permet le développement d'applications autour de OneDoc à l'aide de la bibliothèque d'interfaces programmatique OneDoc API.

### OneDoc Add-ons

En fonction des spécificités professionnelles de l'entreprise ou de l'organisation où il est implanté, le Système de Gestion Documentaire OneDoc peut être enrichi par toute une gamme de modules applicatifs dédiés :

- **OneDoc Workflow (Fujitsu i-Flow)**  
Ce module permet l'implémentation d'un outil décisionnel coopératif de type Workflow, gérant la circulation ordonnancée d'un dossier numérique auprès de personnes ayant une tâche spécifique à accomplir.
- **OneDoc Webfeed**  
Ce module gère l'intégration de documents à partir de centaines de sources d'information sur Internet.
- **OneDoc Exchange**  
Ce module permet la publication et la communication par Echange Electronique de Documents (EDI).

- **OneDoc Notes**

Ce module de publication et de communication est interfacé avec Lotus Notes®.

- **OneDoc Library**

Ce module contient les outils de bibliothéconomie destinés à la gestion d'une bibliothèque ou d'un centre de documentation (traitement des commandes, abonnements, emprunts, etc.).

- **OneDoc Publisher**

Ce module permet la publication sur CD-ROM, sur un serveur Web ou par messagerie électronique d'une partie de la base de données documentaires en vue de sa diffusion auprès d'un réseau d'abonnés.

- **OneDoc COLD**

Cet outil permet l'implémentation de la technologie **COLD** d'archivage sur disque compact (*Computer Output to Laser Disk*) : les documents sous forme de fichiers d'impression sont automatiquement envoyés au système, indexés, puis compressés et stockés afin de les rendre accessibles lors des opérations de recherche (cf une description précise de ce module en page 19 du présent document).

- **OneDoc FCR**

Cet outil permet d'implémenter la technologie **FCR** (*Forms Character Recognition*), destinée à l'acquisition et à la reconnaissance automatique de formulaires remplis de façon manuscrite.

- **OneDoc ICR**

Cet outil permet d'implémenter la technologie **ICR** (*Intelligent Character Recognition*), permettant l'acquisition et la reconnaissance automatique de l'écriture manuscrite.

- **OneDoc Bar Code**

Cet outil permet l'acquisition et la reconnaissance automatique de caractères sous forme de codes barres.

### Solutions Verticales

Pour répondre à des besoins spécifiques du marché, plusieurs solutions sectorielles verticales sont en cours de développement autour du logiciel OneDoc :

- **OneDoc Country/Personality Profile**
- **OneDoc Juridical**
- **OneDoc Technical**
- **OneDoc Photography**
- **OneDoc Audio/Video**

## Choix Technologiques

Lors de l'élaboration de OneDoc, l'équipe de concepteurs logiciels de **OneCor Software** s'est fixé plusieurs objectifs de base:

- **Une Architecture de type Web Serveur :**

Cette architecture est construite exclusivement autour des technologies qui ont fait et continuent de faire le succès d'Internet. OneDoc peut en effet être déployé très rapidement et facilement, puisque son installation se fait uniquement au niveau serveur.

Seul un navigateur Web (de type **Netscape Navigator®** ou **Microsoft Internet Explorer®**) est requis au niveau client. OneDoc peut être utilisé aussi bien en Intranet qu'à travers Internet ou un Extranet.

- **La Technologie Java 2 Entreprise Edition :**

**Java®** est aujourd'hui bien plus qu'un langage de dernière génération orienté objet : c'est une véritable plateforme de développement puissante, moderne, portable et surtout ouverte.

**Sun Microsystems®** a su créer une plateforme non-propriétaire universelle, aujourd'hui adoptée par les plus grands acteurs du monde informatique. OneDoc est construit selon une architecture multi-niveaux évolutive et très sécurisée.

- **Une Plateforme Multi-systèmes:**

La partie serveur de OneDoc est développée entièrement en **Java** : le logiciel peut donc s'exécuter sur n'importe quelle plateforme de système d'exploitation disposant d'une **Machine Virtuelle Java® (Java Virtual Machine)** compatible **Java2®**.

La partie cliente de OneDoc se présente sous la forme d'un logiciel client léger développé en **HTML, XML, DHTML, Javascript®**.

- **Une Plate-forme Indépendante du Serveur Web ou du Serveur d'Application:**

OneDoc Server peut s'exécuter sous n'importe quel serveur Web ou serveur d'application du marché : **Apache®, Microsoft IIS®, Netscape Application Server®, IBM WebSphere®, BEA WebLogic®** etc.

OneDoc est fourni en standard avec le serveur Web et serveur d'application 100% Java **Jetty®**.

- **Un Accès Multi Base de données:**

OneDoc s'appuie le standard d'accès aux bases de données **Java DataBase Connectivity**, et peut donc s'interfacer avec tous les systèmes de gestion de bases de données disposant d'une interface programmatique JDBC.

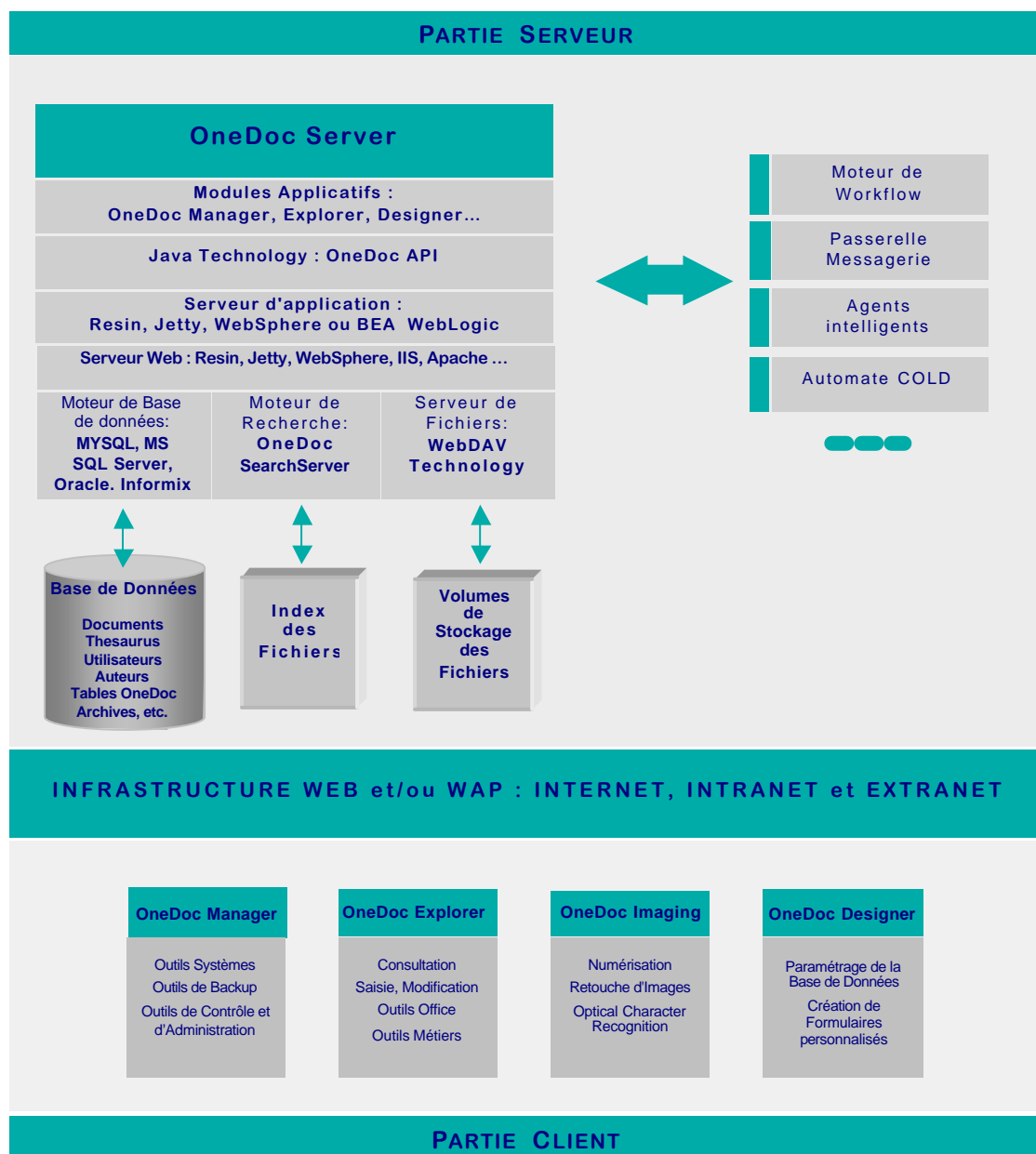
OneDoc est livré en standard avec le Système de Gestion de Bases de Données **MySQL®**, réputé pour sa rapidité, sa capacité à gérer efficacement des bases de données volumineuses et surtout sa popularité sur Internet - MySQL est d'ailleurs devenu librement disponible (open source) .

- **Une Gestion Multilingue et Multi-script:**

OneDoc supporte virtuellement toutes les langues, tant au niveau du contenu des documents qu'au niveau de l'interface utilisateur.

Le moteur d'indexation et de recherche **OneDoc Search Server** est doté d'une fonctionnalité unicode permettant la détection automatique de la langue du texte à indexer.

Le schéma représenté ci-dessous récapitule l'Architecture Technologique du Système de Gestion Documentaire OneDoc.





## Généralités

Bien plus qu'un logiciel de gestion documentaire, OneDoc est une véritable plate-forme de gestion de l'information et des connaissances au sein d'une organisation.

L'objet documentaire tel qu'il est géré par OneDoc est constitué de trois composantes essentielles :

### ▪ Les Propriétés du Document :

Les propriétés du document sont les multiples indications concernant ce document, qui serviront à renseigner l'utilisateur dans ses opérations de recherche :

- Référence,
- Titre,
- Sujet,
- Objet,
- Mots clés,
- Résumé,
- Numéro de version,
- Date de création du document,
- Durée de vie,
- Chemin d'accès et nom du fichier principal,
- Auteur(s),
- Service émetteur,
- Type de document,
- Nature du document,
- Date de publication,
- Type et/ou support de publication,
- Public cible,
- Niveau d'accessibilité, etc.

Les propriétés du document sont structurées dans une base de données. Il est donc possible d'envisager tous les questionnements de l'utilisateur à propos d'un document lors de ses recherches.

OneDoc permet par exemple une recherche portant sur les documents :

- Traitant d'un sujet spécifique,
- Publiés à une date donnée,
- Emis par un service ou un auteur donné,
- S'adressant spécifiquement au profil de l'utilisateur requérant,
- Contenant un mot-clé particulier, etc.

OneDoc gère un nombre illimité d'enregistrement par base de données et peut se déployer sur plusieurs bases de données sous forme de Bibliothèques Documentaires.

OneDoc dispose en standard d'un nombre très élevé de champs qui permettent de renseigner la quasi-totalité des types, formats ou natures de documents.

Un mécanisme de création de champs personnalisés permet d'adapter OneDoc à tous les genres et à toutes les configurations d'applications de gestion documentaire.

### ▪ La Structure du Document :

La structure d'un document est constituée par ses liens hiérarchiques ou liens de rattachement au sein d'une arborescence de documents.

La définition de relations verticales ou horizontales entre plusieurs documents permet de visualiser chacun d'entre eux dans un contexte relationnel particulier.

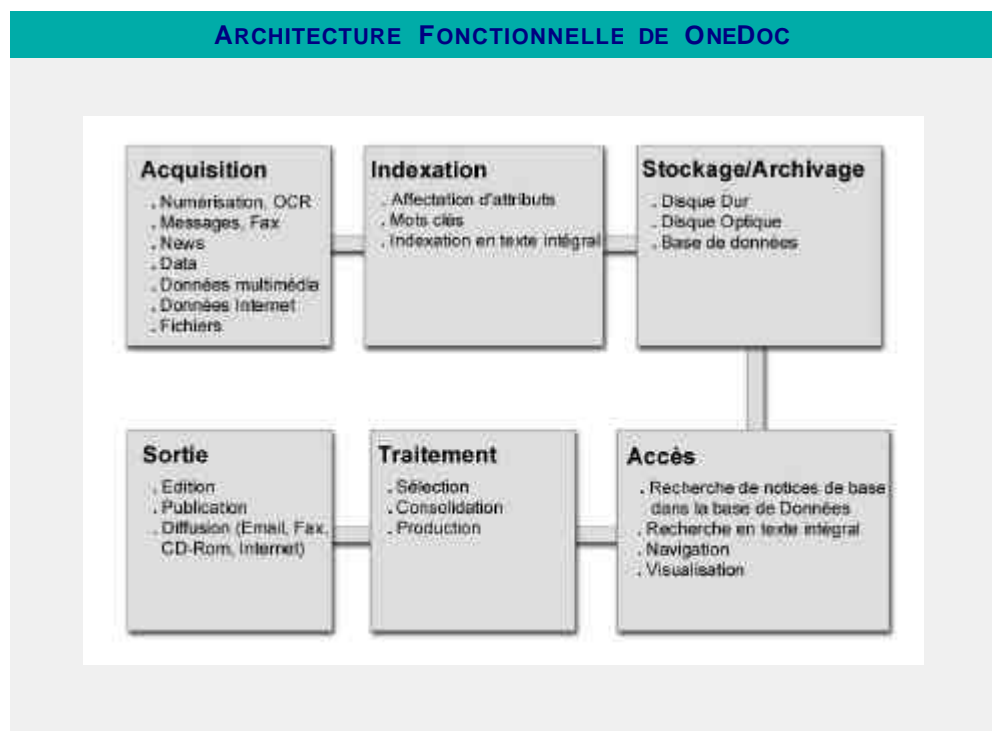
### ▪ Le Contenu du Document :

Le contenu d'un document peut être localisé soit dans la base de donnée, soit dans un fichier externe, soit dans un document physique ou encore dans une page ou un site Internet.

L'indexation en texte intégral du contenu d'un document résidant dans la base de données ou dans un fichier externe permet d'effectuer une recherche sur le contenu de ce document, à l'aide du moteur de recherche en texte intégral.

La fonction de recherche en texte intégral peut être combinée à la recherche multicritères, qui opère sur les champs de la base de données documentaires.

Le schéma représenté ci-dessous illustre l'Architecture Fonctionnelle de OneDoc.



### Acquisition

OneDoc permet plusieurs formes d'intégration des documents dans le système de gestion documentaire : l'acquisition manuelle, automatique ou semi automatique.

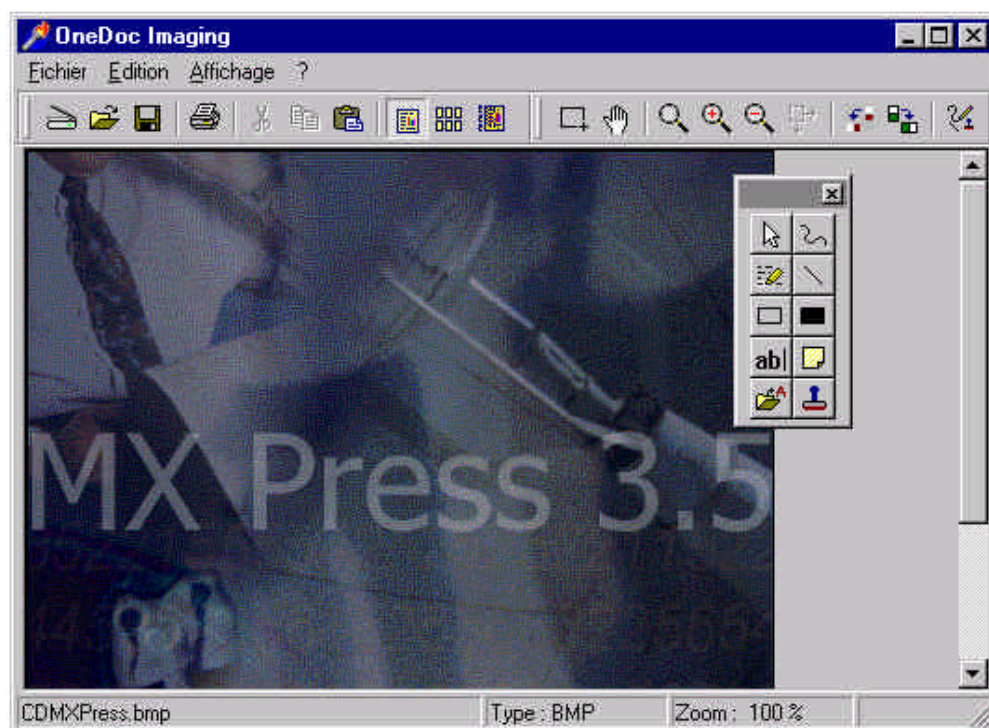
#### ▪ L'Acquisition Manuelle :

OneDoc gère trois modes d'acquisition manuelle des documents qui varient selon le type, le format et la source du document :

- **Numérisation de Documents Papier** au moyen d'un scanner supportant les interfaces **TWAIN**, en utilisant les fonctionnalités du module **OneDoc Imaging** de OneDoc. Tous les formats de papier, du format A5 à A0, sont acceptés. **OneDoc Imaging** permet la retouche des pages numérisées et leur intégration au niveau du serveur documentaire OneDoc Server.
- **Saisie Manuelle des Notices Documentaires** dans l'espace de travail **OneDoc Explorer**, avec ou sans fichiers joints.
- **Intégration des Fichiers Documents** par "enregistrer sous" ou par "glisser-déposer" vers l'espace de travail **OneDoc Space**, présent sur le bureau de la station de travail de l'utilisateur, quel que soit le système d'exploitation utilisé (**Windows®**, **Linux®**, **Solaris®** ou **MacOS®**).

Les fonctionnalités de Gestion Electronique de Documents de OneDoc permettent l'association automatique à des notices documentaires de fichiers externes de toute nature et de tout format (fichiers images, bureautiques, PDF...), sans avoir à sélectionner un à un ces fichiers à partir de la fiche document.

Il suffit de faire glisser les fichiers à archiver sur les documents concernés : OneDoc Space détecte leur présence, effectue leur archivage puis leur suppression. Les éventuels traitements complémentaires sont également effectués : compression de fichiers, extraction d'images (suivant la technologie implémentée), etc.



### ▪ L'Acquisition Automatique :

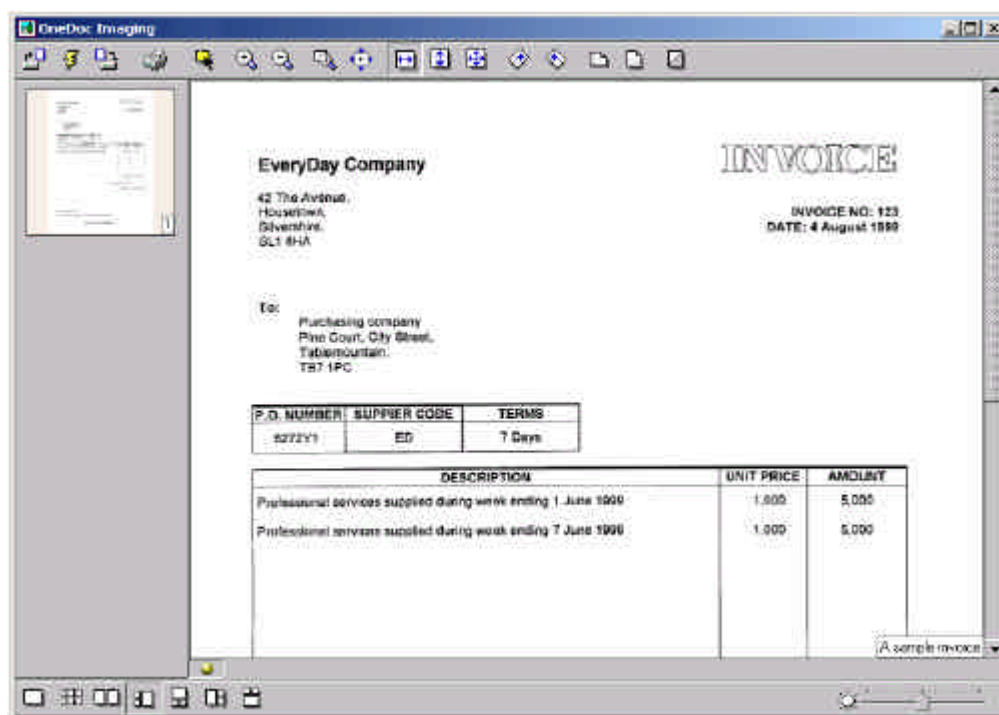
Ce mode d'acquisition permet l'intégration dans la base de données documentaires de documents externes ou importés :

- Réception de flux permanents de documents, tels que les fils de diffusion de dépêches d'agences de presse, les *webfeeds* (sources d'informations générées par l'indexation dynamique du contenu de milliers de pages Web), les passerelles de messagerie électronique, les passerelles fax ou télex, etc.
- Import automatisé de notices sans utiliser la fonction de téléchargement. Cette fonctionnalité est activée par la création d'un dossier *Boîte aux lettres* dans lequel il suffit de glisser le format de téléchargement à utiliser et le fichier à importer (au format texte ".txt").

Le composant **OneDocSpace** détecte la présence d'un nouveau fichier à traiter et procède à son importation, en respectant les critères de détection de doublons et les règles d'intégration choisies.

Il est ainsi possible de faire régulièrement des interrogations de bases de données externes et de télécharger automatiquement les notices intéressantes.

- Extraction automatique de documents archivés à partir de fichiers d'impression stockés sur CD-Rom réinscriptible ou sur disque optique numérique au moyen d'un automate d'archivage **COLD** (Computer Output to Laser Disc).



### ▪ L'Acquisition Semi-Automatique :

Ce mode d'acquisition permet l'import ou le téléchargement de fichiers contenant les notices documentaires au format normalisé **UNIMARC®** ou dans un autre format prédéfini, ainsi que l'import des fichiers qui sont attachés à ces notices.

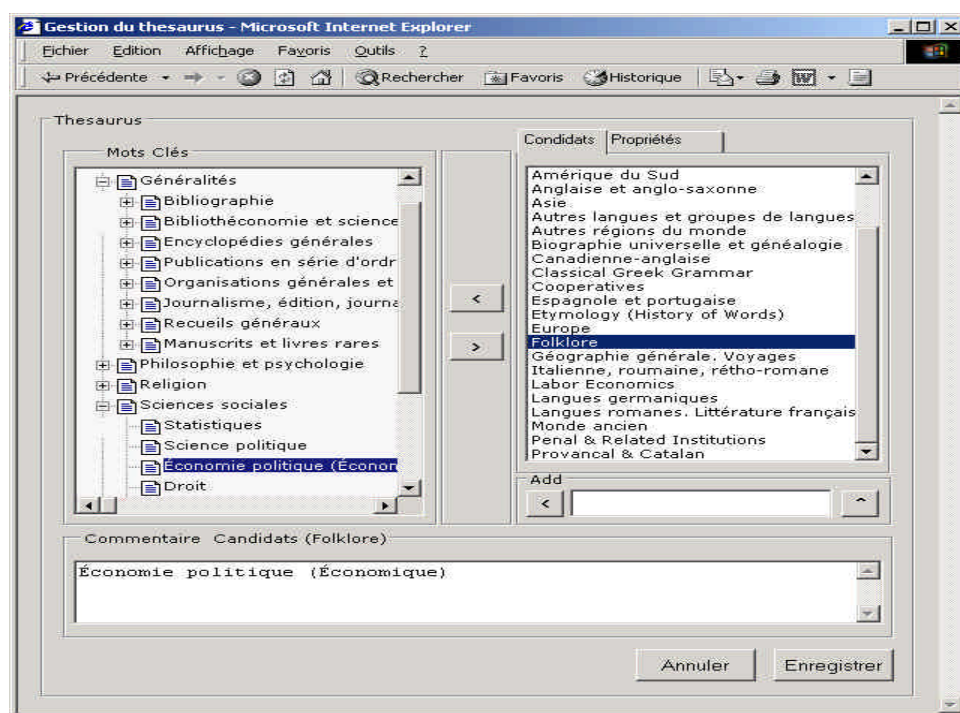
## Indexation

Le logiciel OneDoc gère plusieurs types d'indexation : l'**Indexation Documentaire**, l'**Indexation par Mots-clés** et l'**Indexation en Texte Intégral**.

### ■ L'Indexation Documentaire:

L'indexation documentaire s'effectue en réalisant les opérations suivantes:

- Alimentation des champs représentant les propriétés qui renseignent sur le document,
- Affectation des mots clés liés hiérarchiquement dans une relation horizontale indépendante de leurs positions et de leurs relations au sein du thésaurus,
- Création des liens hypertextes simples (point à point) ou multiples entre les différents documents de la base de données.



### ■ L'Indexation par Mots-Clés:

L'indexation par mots clés s'appuie sur des termes spécifiques appelés **Descripteurs**, regroupés dans un **Thésaurus**. Le logiciel OneDoc gère plusieurs thésaurus de façon multilingue, c'est à dire qu'il permet d'effectuer la correspondance entre des descripteurs de langues différentes.

L'efficacité de toute recherche documentaire dépend complètement de la pertinence du thésaurus utilisé et de son adéquation avec le domaine couvert. La construction du thésaurus propre à une organisation constitue donc une étape importante et délicate qu'il convient de traiter avec la plus grande attention.

Dans cette optique, le système OneDoc permet la conception normalisée (**ISO 2709**) d'un thésaurus multilingue et propose différentes manières de procéder à la création du thésaurus :

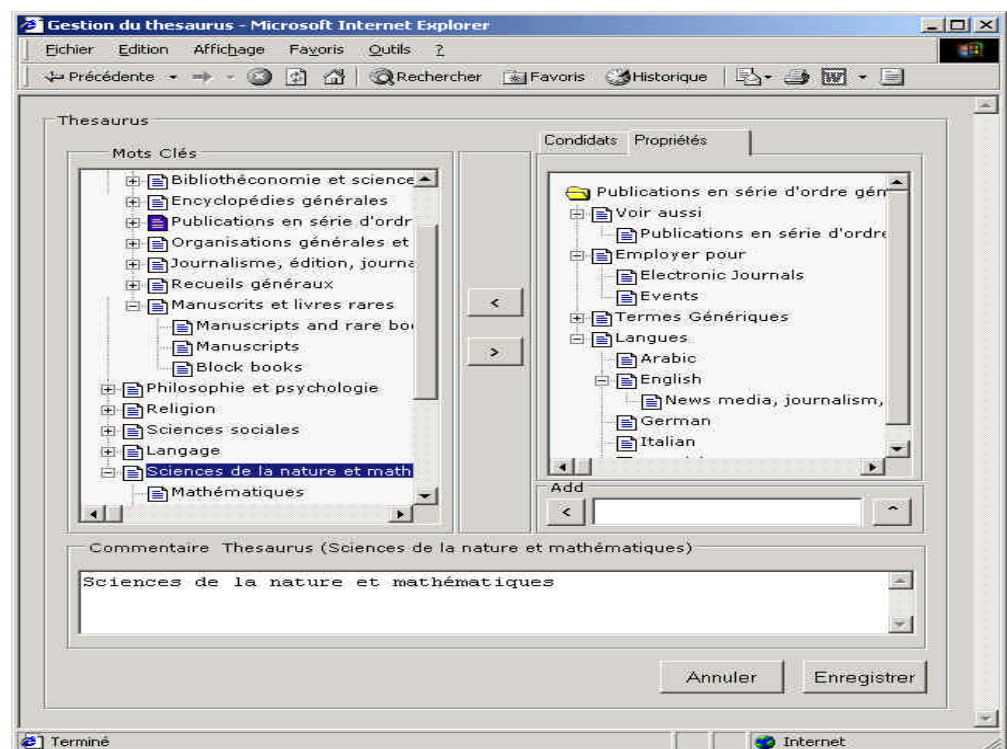
- Si le thésaurus existe déjà sur papier, il est possible de procéder à sa saisie directe dans OneDoc ;
- Si le thésaurus est partiellement ou totalement saisi avec un logiciel de traitement de texte (par exemple Word®), un utilitaire de conversion permet son importation directe dans OneDoc ;

- Si le thésaurus est partiellement créé : il est possible d'effectuer la saisie des premiers niveaux hiérarchiques puis de créer au fur et à mesure l'arborescence des candidats descripteurs ;
- Si le thésaurus n'est pas constitué : il est possible de réaliser la création de descripteurs candidats puis la structuration a posteriori de ces descripteurs candidats ;

OneDoc permet également l'importation d'un thésaurus informatique externe.

De plus, une panoplie d'outils rend possibles le suivi du thésaurus et la réalisation des mises à jour si nécessaire :

- Edition de listes de descripteurs avec compteur (fréquence d'utilisation à l'indexation) ;
- Suppression d'un terme du thésaurus sur les fiches indexées ;
- Remplacement d'un terme du thésaurus par un autre ;
- Comptage du nombre des utilisations de chaque descripteur lors des procédures de recherche.



OneDoc gère également plusieurs types de relations entre descripteurs:

- **Relations Hiérarchiques** entre un descripteur et son terme spécifique : par exemple, « *clavier* » est un terme spécifique du descripteur « *matériel informatique* » ;
- **Relations d'Equivalence** entre un descripteur et ses synonymes, par exemple « *éducation* » et « *enseignement* » ;
- **Relations d'Association** entre deux termes voisins mais appartenant à des domaines différents, par exemple « *enseignement* » et « *formation professionnelle* ».



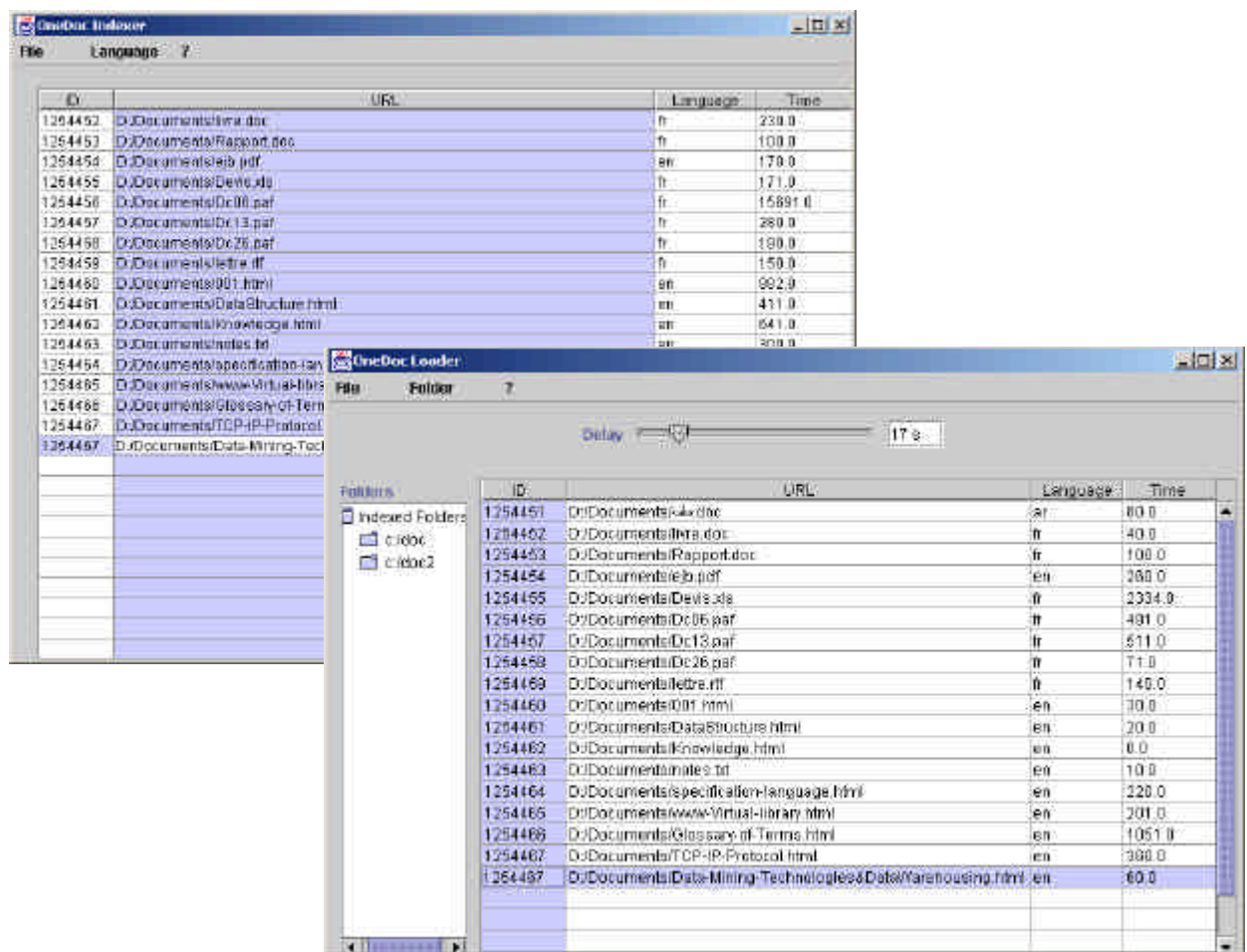
### ▪ L'Indexation en Texte Intégral :

L'indexation en texte intégral consiste à analyser l'ensemble des contenus rattachés à un document et à les indexer à l'aide du moteur d'indexation de OneDoc, qui permet une indexation multilingue du texte et une recherche sur le contenu du texte. Le moteur d'indexation de OneDoc, **OneDoc Index Server**, constitue l'une des composantes du package **OneDoc Search Server**, le serveur d'indexation et de recherche en texte intégral de OneDoc qui a été développé à l'aide du toolkit **Lucene®**.

Le moteur d'indexation de OneDoc est dit **multiprocess**, car il est capable d'exécuter simultanément plusieurs instances du processus d'indexation. Chaque processus réalise l'indexation d'une collection de documents et génère les index correspondants sur le volume de stockage.

Cette technique permet de faire fonctionner plusieurs disques en même temps, à la fois pour la lecture des documents à indexer et pour l'écriture des index générés. Ces capacités de **multithreading** (exécution simultanée de séquences d'instructions multiples par le noyau dans une application multiprocess) permettent à l'indexation de s'effectuer sur de gros volumes de données.

**OneDoc Index Server** peut également se déployer sur plusieurs serveurs, augmentant de ce fait la vitesse d'indexation des collections de documents.



### Stockage

Chaque document est constitué d'une composante "*propriétés*" : ces propriétés sont stockées dans les champs des tables de la base de données. OneDoc permet de s'interfacer avec tous les Systèmes de Gestion de Bases de Données disposant d'une interface programmatique au standard **Java DataBase Connectivity (JDBC)**.

La composante "*contenu externe*", qui réside dans un ou plusieurs fichiers externes rattachés aux documents, est stockée dans des volumes de stockage destinés à cet effet, qui peuvent être des volumes de disques durs, des Jukebox de disques WORM, des CD-ROM via un graveur, des DVD, etc.

Les fichiers sont maintenus dans leur format natif et peuvent être compressés à l'aide de la technologie **Zip**. Certains fichiers, tels que les images, peuvent être compressés au format **JPEG** ou **TIFF**.

### Recherche

Le module de recherche de OneDoc permet de faire porter la recherche soit sur les données structurées présentes dans la base de données, soit sur le contenu des textes associés à un document.

Il est également possible de combiner ces deux types de recherches. OneDoc propose plusieurs modes de recherche : chacun de ces modes de recherche est disponible au moyen d'un formulaire prévu à cet effet:

- **Recherche rapide :**

Ce mode de recherche est particulièrement adapté aux consultations rapides. L'utilisateur dispose d'une zone de texte libre où il peut saisir ses critères de recherche. Un ensemble de boîtes à cocher lui permettent de sélectionner les champs sur lesquels va porter la recherche.

- **Recherche étendue :**

Ce mode de recherche permet de questionner la base documentaire de manière plus exhaustive, en combinant ensemble plusieurs critères de recherche au moyen des opérateurs booléens Et, Ou, Sauf.

- **Recherche avancée :**

Ce mode de recherche convient plus précisément à une utilisation professionnelle : les recherches peuvent être composées élément par élément, c'est à dire qu'elles peuvent porter sur chacun des champs de la base de données et ce de façon séparée.

Chaque recherche unitaire donne lieu à la production d'une liste de résultats, indiquant le nombre de documents identifiés. Un simple clic sur le bouton "**Afficher**" permet de visualiser les documents de la liste.

Il existe également un mode de recherche combinatoire qui permet de combiner à volonté les critères de recherches unitaires déjà effectuées par les opérateurs booléens (Et, Ou, Sauf). Ce mode de recherche permet de réaliser de manière très souple des recherches d'une grande complexité.

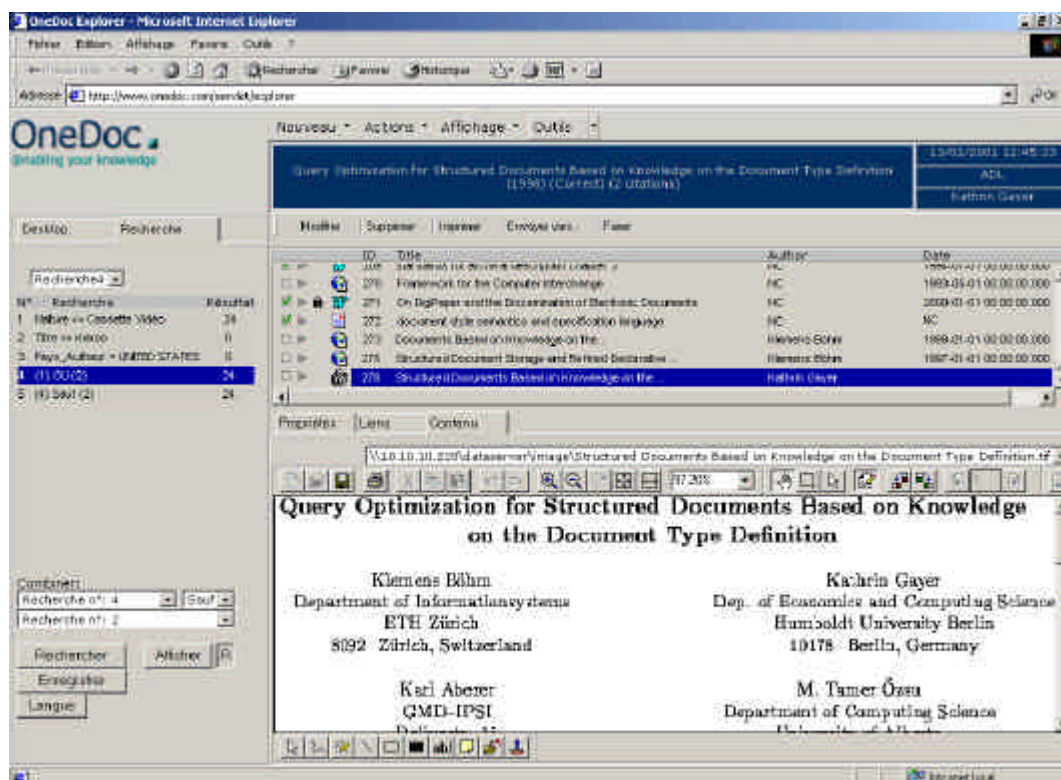
Il est également possible à l'utilisateur d'enregistrer les stratégies de recherches déjà effectuées, de manière à les ré-exploiter ultérieurement au niveau de l'arbre de navigation ou du formulaire de recherche lui-même.



## ■ Recherche personnalisée :

La recherche personnalisée est une forme de Recherche par l'exemple ( *Query By Example*), où chaque utilisateur peut personnaliser et exploiter les formulaires de recherche spécifiques qu'il a créés en fonction de ses besoins.

Un ensemble de formulaire QBE, livrés en standard avec OneDoc, est accessible à tous les utilisateurs.



## Navigation

OneDoc dispose d'un mode de navigation à travers une zone de liste hiérarchique, qui permet de constituer des sélections de documents par un simple clic sur l'un de ses éléments. Ce mode de navigation est très puissant : il permet d'avoir une vision globale et multiforme sur le patrimoine de connaissances géré par l'entreprise.

Outre les éléments de navigation partagés ou publics, chaque utilisateur possède la faculté de créer ses propres objets de navigations organisés et classés dans des dossiers. L'utilisateur peut par exemple disposer des éléments suivants :

- **Objet "Thesaurus"**

L'objet Thésaurus est constitué de toute une liste hiérarchique contenant les descripteurs qui le composent. En cliquant sur chacun des descripteurs, l'utilisateur déclenche automatiquement une recherche dans la base de données documentaires pour afficher dans la zone de résultat la sélection des documents trouvés.

- **Objet "Catégories"**

L'objet Catégories permet également de naviguer à travers la base documentaire par catégorie.

- **Objets Recherches**

Les Objets Recherches sont constitués des recherches préalablement enregistrées à partir d'un formulaire de recherche.

- **Objets Casiers**

Les Objets Casiers contiennent des documents déposés manuellement par l'utilisateur.

La zone de navigation est ainsi constituée de trois espaces de travail principaux :

- **Portal Space:**

Cette zone contient les liens vers l'ensemble des portails Web fréquemment accédés par l'utilisateur (signets), par exemple **MyYahoo®**, **MyLycos®** ou encore **OnePortal**, le portail d'informations d'entreprise édité par OneCor.

- **Knowledge Space:**

Cette zone est également constituée de trois dossiers. Ainsi, l'organisation particulière des objets de navigation permet de tracer une véritable cartographie de tous les gisements de connaissances existant au sein de l'entreprise (connaissances personnelles, partagées ou publiques) et de traduire de façon claire l'organisation interne de l'entreprise.

- **Personnal Folder:** Ce dossier contient l'ensemble des objets de navigation personnels de l'utilisateur.

- **Public Folder:** Ce dossier contient les objets de navigation accessibles à tous les utilisateurs.

- **Shared Folder:** Ce dossier contient l'ensemble des objets de navigation partagés, c'est à dire spécifiques à un groupe d'utilisateurs donné.

- **Application Space:**

Cette zone contient des éléments applicatifs divers (accès à un serveur de messagerie électronique, forums, chats, applications Web, etc.).

## Consultation

La consultation des documents s'effectue à partir de la liste de documents résultant d'une opération de recherche ou de la sélection d'un élément de navigation.

La liste des documents permet à l'utilisateur de réaliser diverses manipulations :

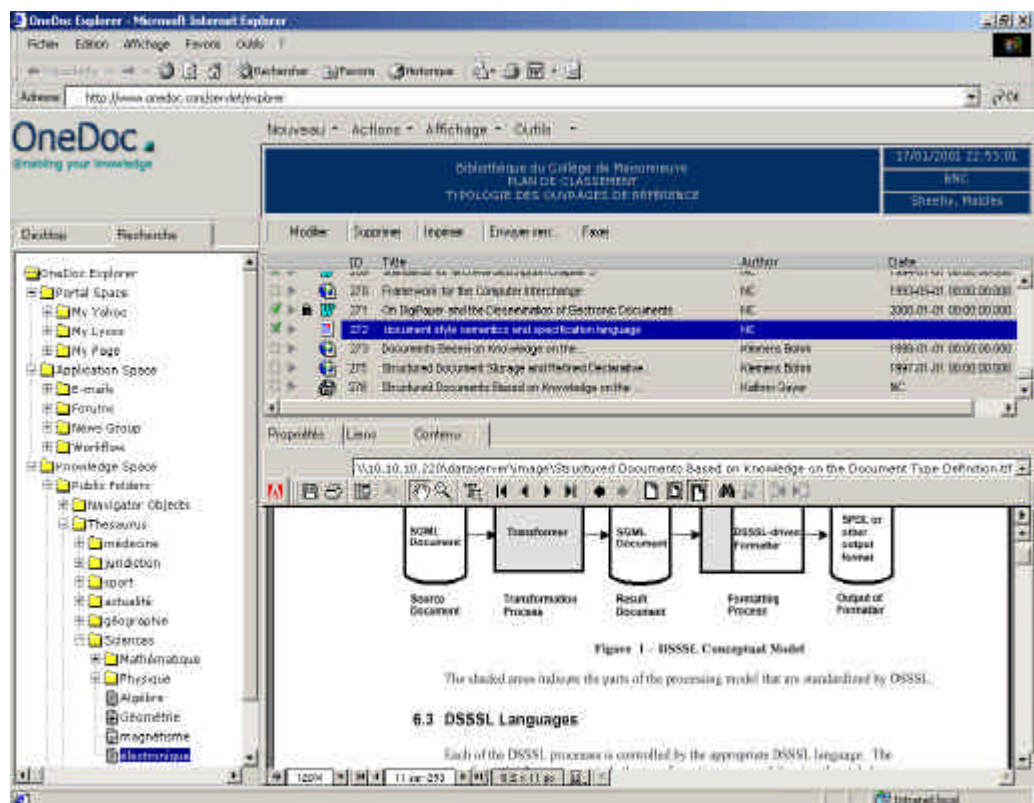
- **Tri, Marquage et Déploiement des documents liés**

Il est également possible de réaliser des actions spécifiques pour chaque document.

- **Modification, Suppression, Impression, Envoi vers...**

Un simple clic sur le nom d'un document affiche à la fois ses propriétés, ses liens et ses contenus. Le contenu du document peut être visualisé à l'aide de l'outil **OneDoc Viewer** ou si nécessaire à l'aide d'autres outils (*plug-ins, ActiveX ou applets*).

Le document est visualisé soit directement depuis l'explorateur, soit dans une fenêtre externe.



### OneDoc COLD      **La Technologie COLD : Conversion de Fichiers d'Impression en Archives Accessibles par des Opérations de Recherche**

Le module **OneDoc COLD** ajoute une nouvelle fonctionnalité à la plate-forme de gestion documentaire et de gestion de connaissance OneDoc.

Avec OneDoc COLD, les fichiers d'impression (fichiers spool) sont automatiquement traités et rendus accessibles aux opérations de recherche.

#### **Rappel**

Les ordinateurs centraux et départementaux produisent en général de gros volumes de données diverses : statistiques, balances, factures, etc. Dans les entreprises de grande taille, ce volume peut atteindre plusieurs centaines de milliers de pages par jour. Ces données sont habituellement envoyées vers une imprimante départementale pour y être imprimées : si le volume quotidien de fichiers d'impression est élevé, la quantité de papier produite peut être énorme.

L'étape suivante a consisté à envoyer le résultat de ces impressions sur des microfiches plutôt que vers une imprimante. Ce fut l'avènement de la technologie **COM (Computer Output on Microfilm)**, qui présentait plusieurs avantages : réduction des dépenses de transport, simplification de la distribution plus facile, réduction de l'espace de rangement nécessaire (plus de papier).

L'utilisation généralisée du disque optique l'a progressivement imposé comme support de stockage (disque laser WORM par exemple). Le processus de stockage lié à ce type de support est appelé **COLD (Computer Output on Laser Disc)**, et le type de disque laser le plus fréquemment utilisé est le CD-ROM.

Les fichiers produits par les ordinateurs centraux et départementaux sont des **fichiers d'impression** ou fichiers spool. Ce sont en général des fichiers très volumineux - plusieurs centaines de méga-octets par fichier -.

Le stockage sur disque de ces énormes fichiers les rend très difficiles à manipuler. Chaque fichier spool contient un en-tête décrivant le contenu du fichier : c'est ce champs qui est utilisé pour retrouver l'information, un numéro de client par exemple.

#### **Description**

Le module **OneDoc COLD** a été développé pour assurer l'automatisation du processus de stockage des fichiers d'impression sur disques durs ou disques compacts ou pour permettre leur publication sur Internet.

Ce module permet également de structurer l'information et d'effectuer des recherches en texte intégral. Pour l'utilisateur, l'avantage est immédiat : les gros fichiers sont scindés en fichiers de petite taille beaucoup plus faciles à manipuler et contenant chacun un champs ou en-tête descriptive.

Ces petits fichiers sont indexés en texte intégral et deviennent donc accessibles aux procédures de recherche. L'utilisateur peut ainsi effectuer des recherches sur chacun des mots contenus dans ces fichiers ainsi que sur les champs d'en-tête des fichiers.

OneDoc COLD détecte automatiquement la présence de nouveaux fichiers d'impression dans un répertoire, répartit ces fichiers en plusieurs fichiers séparés, leur ajoute des champs structurés et indexe rapidement la totalité de l'information qu'ils contiennent. Une fois effectués les contrôles d'intégrité et de qualité, l'index peut être publié sur disque compact ou sur Internet.

Le module OneDoc COLD peut également être utilisé pour diviser en fichiers séparés des fichiers texte, des archives de messagerie électronique, des informations provenant d'Internet ou toute autre information stockée dans un fichier texte.

Avec le module de publication sur disque compact **OneDoc CD Publisher** intégré au logiciel OneDoc, OneDoc COLD supporte la diffusion d'index et de gros volumes de données sur disques multiples, qui peuvent ensuite être accessibles à travers un juke-box ou une tour de disques compacts.

Il est également possible de publier le moteur de recherche de OneDoc sur le disque compact pour créer un disque autonome.

Le logiciel OneDoc permet de partager des données identiques à travers le réseau Internet et dans un Intranet d'entreprise. La flexibilité de OneDoc permet à l'utilisateur de personnaliser entièrement l'interface de recherche du logiciel (par exemple en limitant les procédures de recherche à certains champs) et de concevoir un masque personnalisé qui corresponde au modèle original du document.

### Fonctionnement

Le processus d'archivage par OneDoc COLD des fichiers d'impression sur disque compact se compose des sept étapes suivantes :

- **Identification du Fichier d'Impression :**

les fichiers prêts à être traités sont envoyés dans un répertoire spécifique, où OneDoc COLD vérifie l'arrivée de nouveaux fichiers. Dès qu'un nouveau fichier est détecté, le programme commence à le traiter.

- **Division du Fichier d'Impression :**

Les fichiers d'impression étant en général très volumineux, il est plus aisé de les scinder avant de poursuivre le déroulement du processus. La méthode de division des fichiers peut être contrôlée avec un script.

Pendant le déroulement du processus de division des fichiers, OneDoc COLD crée automatiquement une arborescence de sous-répertoires en affectant un certain nombre de fichiers par répertoire, afin d'optimiser la vitesse d'accès aux fichiers. La valeur d'affectation par défaut est de 256 fichiers par répertoire.

- **Conversion des Jeux de Caractères :**

Pendant le déroulement du processus de division des fichiers, certaines chaînes de caractères présentes dans les fichiers texte peuvent être automatiquement remplacées par d'autres chaînes de caractères choisies par l'utilisateur.

- **Ajout de la Structure :**

L'administrateur du système doit écrire un fichier script permettant le contrôle du processus de division (OneDoc COLD propose un fichier script par défaut). Il est possible de spécifier dans le scénario le nom de l'index à produire et également les critères de division des fichiers d'impression : division en fonction d'une longueur fixe ou lors de chaque occurrence d'une certaine chaîne de caractères.

L'ajout de champs et de concepts dans l'index peut également être spécifié dans le scénario : l'en-tête de chaque fichier d'impression contient en général ces données. Les champs supplémentaires utilisent le format XML, à l'instar des autres produits de la plate-forme logicielle OneDoc.

- **Création de l'Index :**

Après la division des fichiers et leur répartition dans les sous-répertoires et après l'ajout de la structure, OneDoc COLD procède à la construction de l'index, en fonction des spécifications contenues dans le fichier script. Le programme affecte un nom unique à chaque index d'après une convention de nommage automatique.

- **Additif :**

En procédant à la numérisation des états papiers imprimés par les fichiers d'impression, il est possible d'utiliser les images ainsi obtenues comme un additif permettant de visualiser les fichiers séparés dans leur format original à l'aide du navigateur Internet.

- **Gravure du CD ou publication sur le web :**

Le module de publication sur disque compact de OneDoc, **OneDoc CD Publisher**, permet à l'utilisateur de créer son propre disque compact indexé complété par le moteur de recherche de OneDoc, ou de publier avec OneDoc Server ces informations sur Internet ou sur l'Intranet de l'entreprise.

Les disques compacts créés par le module **OneDoc CD Publisher** sont compatibles avec la plupart des juke-box existants.

En plus des opérations de recherche portant sur les champs, les utilisateurs peuvent ainsi effectuer des recherches dans le texte intégral du document, par exemple rechercher un nom de client ou une ville et exploiter au maximum l'ensemble des fonctionnalités de recherche de OneDoc.

Si les données sont publiées sur Internet avec OneDoc Server, il est possible de numériser le document original et de l'employer comme fond d'écran pour la visualisation des fichiers divisés, ce qui donne à l'utilisateur l'impression d'avoir sous les yeux l'original du document.

## Architecture

L'architecture de développement de OneDoc est une architecture multi-niveaux :

- **Partie Serveur :**

La Partie Serveur est développée selon les principes de l'Architecture **MVC (Model-View-Controller)** basée sur la technologie **Java2 Enterprise Edition®**: **Servlets, JSP, JavaBeans** et **Enterprise JavaBeans**.

- **Partie Client**

La Partie Client est un programme léger accessible via un navigateur du marché comme **Netscape®** ou **Internet Explorer®**, et développé à l'aide des langages et méta langages **Javascript,HTML, DHTML, XML, XHTML** .

Le langage de développement retenu pour le développement de OneDoc est **Java®**.

Les standards de Systèmes de Gestion de Bases de Données retenus sont **SQL (Structured Query Language), JDBC (Java DataBase Connectivity)**.

Le standard de connectivité réseau choisi est **TCP/IP**.

## Systèmes d'Exploitation

Les systèmes d'exploitation supportés par OneDoc sont les suivants :

- **Microsoft Windows 95, 98, ME, NT, 2000®**
- **Linux®**
- **Sun Solaris®**
- **IBM AIX®**
- **MacOS®.**

## SGBD

Les Systèmes de Gestion de Bases de Données supportés par OneDoc sont les suivants:

- **MYSQL®**
- **MS SQL Server®**
- **Sybase®**
- **Oracle®**
- **Informix®**

## Standards Technologiques

Les standards technologiques supportés par OneDoc en matière de télécommunications et de langages de développement sont les suivants :

- **TCP**
- **HTTP, HTTPS, SSL, WEBDAV**
- **HTML, CSS**
- **JAVA2EE®**
- **WAP, WML**
- **XML, XSL**
- **SOAP, JXTA**
- **LDAP, MIME**
- **JavaScript, DHTML**
- **Unicode**

## Standards Documentaires

Les principaux standards documentaires supportés par OneDoc sont les suivants :

- **ISO** (pour le thésaurus, les auteurs, les catégories)
- **ODMA** etc.