

19) RÉPUBLIQUE FRANÇAISE  
INSTITUT NATIONAL  
DE LA PROPRIÉTÉ INDUSTRIELLE  
PARIS

11) N° de publication :  
(à n'utiliser que pour les  
commandes de reproduction)

2 852 712

21) N° d'enregistrement national : 03 03222

51) Int Cl<sup>7</sup> : G 06 F 17/24

12)

## DEMANDE DE BREVET D'INVENTION

A1

22) Date de dépôt : 17.03.03.

30) Priorité :

43) Date de mise à la disposition du public de la demande : 24.09.04 Bulletin 04/39.

56) Liste des documents cités dans le rapport de recherche préliminaire : *Se reporter à la fin du présent fascicule*

60) Références à d'autres documents nationaux apparentés : Division demandée le 17/03/04 bénéficiant de la date de dépôt du 17/03/03 de la demande initiale n° 03 03222.

71) Demandeur(s) : ROCHER GILLES PHILIPPE — FR.

72) Inventeur(s) : ROCHER GILLES PHILIPPE.

73) Titulaire(s) :

74) Mandataire(s) : ROCHER GILLES.

54) **SYSTEME DE TRAITEMENT INFORMATIQUE PERMETTANT A DES UTILISATEURS D'UN RESEAU DE COMMUNICATION, NOTAMMENT DU TYPE INTERNET, DE GENERER DES DOCUMENTS FONCTIONNELS A PARTIR D'INFORMATIONS TECHNIQUES.**

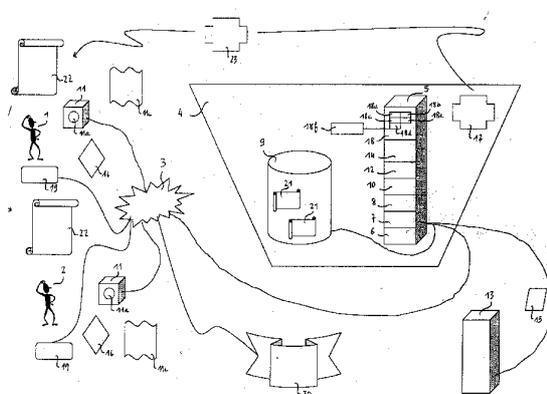
57) Système de traitement informatique permettant à des utilisateurs (1, 2) d'un réseau de communication (3) de générer des documents (23) dans un système de paiement sécurisé comprenant une plate-forme distante (4) comportant :

- des premiers moyens de calcul (6) pour calculer une base tarifaire,
- des seconds moyens de calcul (7) pour calculer une fraction variable tarifaire,
- des moyens de comparaison (8) pour comparer la base tarifaire avec la fraction variable et calculer une pondération.

La plate-forme distante (4) comprend une base de données (9). La plate-forme distante (4) est connectée à des équipements informatiques (11). Le système comprend en outre :

- des premiers moyens de transmission (14) pour transmettre des informations de paiement (15),
- des seconds moyens de transmission (16) pour transmettre des informations techniques à la plate-forme distante (4),
- des moyens d'extraction,
- des moyens d'assemblage, pour générer un document

final (22).



FR 2 852 712 - A1



**Système de traitement informatique permettant à des utilisateurs d'un réseau de communication, notamment du type Internet, de générer des documents fonctionnels à partir d'informations techniques.**

5 La présente invention concerne le domaine de la sémantique dans ses applications à la génération de documents comportant une fonctionnalité prédéterminée. Depuis le milieu des années 1980, la sémantique, sans s'éloigner de la rigueur méthodologique fondatrice (voir notamment les travaux de Frege et Montague) a élaboré des théories et des formalismes destinés à assurer la flexibilité et la couverture empirique requises pour un  
10 traitement fin des données linguistiques. Des éléments initiateurs très significatifs de ce développement sont la généralisation du modèle de la quantification, l'émergence des théories dynamiques et des re-délimitations des domaines respectifs de la sémantique et de la pragmatique.

15 La sémantique dite formelle inclut des étapes descriptive et analytique elles-mêmes suivies d'une étape de modélisation, qui vise à valider les prédictions fondées sur les observations et, plus fondamentalement, à pénétrer les mécanismes abstraits de construction du sens. Plusieurs études, connues de l'art antérieur, montrent la valeur ajoutée de la modélisation dans une démarche explicative et donc prédictive. Des modèles  
20 algorithmiques nouveaux permettent désormais de fusionner des représentations structurées (éventuellement révisables) et non structurées d'une même source d'information technique (un document). Ces nouveaux modèles cherchent à harmoniser des concepts développés dans les domaines de l'indexation automatique des données multimédia, du  
25 traitement du langage naturel et des bases de données relationnelles, objets ou déductives, dans le but d'améliorer sensiblement la qualité des accès à l'information archivée, moyennant un effort de structuration des données raisonnable, capitalisable et révisable. Ces recherches ont notamment donné lieu à la réalisation d'outils d'indexation automatique, de recherche

et de filtrage d'information multimédia, disponibles au format XML. Des logiciels dédiés au traitement des données, de type séries temporelles ou séquentielles, sont apparus. La recherche par similarité dans les banques de séries temporelles ou séquentielles de grande taille joue un rôle croissant dans nombre d'applications qui relèvent de la recherche d'information multimédia, du "data mining" ou de la classification automatique. En particulier, on peut citer l'analyse de sous-séquences génomiques, la comparaison de données pathologiques évolutives, la recherche de similarité dans les cours des valeurs boursières, la recherche de similarité dans les partitions musicales et/ou bandes sonores utilisée notamment dans la gestion des droits d'auteurs. Or dans la plupart des cas de réalisations logicielles, l'objectif fonctionnel poursuivi n'est que très rarement atteint du fait d'une confusion, ou assimilation fréquente, entre bases de données dynamiques et bibliothèques virtuelles accessibles notamment via un réseau de type Internet. En effet, les bibliothèques virtuelles ou bibliothèques numériques doivent permettre de retrouver en ligne toutes les facettes des bibliothèques classiques (réservoir de documents, lieu de confrontation du lecteur avec le multimédia, lieu de conservation de documents, lieu d'organisation des connaissances par la classification et le catalogage, aide bibliographique). De ce fait, la mise en place de bibliothèques virtuelles implique plusieurs domaines de recherche : la numérisation des documents, la classification et l'indexation des documents pour améliorer leurs recherches, l'accès à distance aux documents avec l'utilisation d'architectures informatiques adaptées, les problèmes de droits d'auteurs pour la diffusion de documents électroniques ainsi que l'aspect social d'un tel système. Cette mise en place suppose donc la modélisation d'architectures informatiques qui constituent un support pour accueillir les fonctionnalités nécessaires à la mise en œuvre des bibliothèques virtuelles. Cette modélisation s'articule le plus souvent

sur les axes de réalisation décrits ci-après. Tout d'abord une élaboration d'un modèle de représentation logique et physique de documents basé sur les normes internationales (en particulier HTML et XML). On peut ainsi spécifier aisément des modèles de documents appelés DTD (Document Type Définition) permettant de fournir une structure d'accueil pour l'instanciation de documents quel que soit leur type (livre, périodique). Ce modèle prend en compte la composante multimédia de certains types de documents (image, son, vidéo). Ces données multimédias sont très riches en informations et nécessitent des mécanismes de description élaborés. Cette phase de modélisation constitue une étape indispensable car elle influence de manière non négligeable l'étape suivante à savoir le processus d'indexation. Cette étape d'indexation permet de capturer les informations contenues dans les documents, dans le but de faciliter la phase de recherche d'informations. La méthode d'indexation est destinée à extraire deux types d'informations :

- les méta données (auteur, titre, résumé, format, date de création), notamment sous la norme du Dublin Core,
- l'information liée au contenu sémantique du document en se basant sur les graphes conceptuels.

Un modèle utilisateur est ensuite développé pour permettre d'exploiter le contenu des bibliothèques. Le modèle utilisateur intègre les aspects comportementaux et cognitifs. Il est utilisé par les modules de recherche et d'extraction d'information pertinente. En fonction de spécifications inhérentes aux cahiers des charges, des modèles et des outils de collaboration, aptes à permettre un accès transparent à des documents gérés par un ensemble de bibliothèques, sont mis en place. Leur gestion est alors assurée comme celle un espace documentaire virtuel unique. On mesure donc ainsi les limites d'un tel type de modèle en ce sens qu'il ne peut générer que des documents à la fonctionnalité statique selon un

processus d'élaboration que l'on pourrait qualifier d'aveugle car l'espace documentaire virtuel et les moyens d'y naviguer, on vient de le voir, sont rendus involontairement déstructurés.

5 La présente invention propose de résoudre ce problème grâce à la mise en place et à l'administration d'un système de traitement informatique permettant à des utilisateurs d'un réseau de communication, notamment du type Internet, de générer des documents fonctionnels à partir d'informations techniques. La présente invention permet de rendre compatibles une approche quasi ontologique du traitement sémantique d'une part et une ergonomie d'utilisation correspondant aux standards de navigabilité de la plupart des sites Internet marchands d'autre part. Les utilisateurs n'ont ainsi pas besoin de posséder des qualifications particulières ou des connaissances particulières dans le domaine du traitement sémantique pour obtenir des documents destinés à faire sens vis-à-vis des spécialistes du domaine. Pour cela, le système de traitement informatique est utilisé dans un système de paiement sécurisé destiné à permettre aux utilisateurs du réseau de communication, notamment du type Internet, d'effectuer des paiements destinés à l'acquisition des documents fonctionnels proposés en vente par une plate-forme distante des utilisateurs du réseau de communication. La plate-forme distante du système de paiement sécurisé comprend un serveur. Le serveur comporte des premiers moyens de calcul pour calculer, en fonction d'une série de paramètres fixes, une base tarifaire. Le serveur comporte en outre des seconds moyens de calcul pour calculer, en fonction d'une série de paramètres variables, une fraction variable tarifaire. Le serveur comporte en outre des moyens de comparaison pour comparer la base tarifaire avec la fraction variable et calculer une pondération en fonction de la comparaison ainsi effectuée. La plate-forme distante comprend en outre une base de données connectée audit serveur. La plate-forme distante comprend en

10

15

20

25

outre des moyens de mémorisation pour mémoriser, dans la base de données, la série de paramètres fixes et la série de paramètres variables et la base tarifaire associée à la pondération. La plate-forme est connectée, via le réseau de communication informatique, à des équipements informatiques.

5 Les équipements informatiques comportent des moyens de visualisation sur lesquels les utilisateurs peuvent lire des informations diffusées par la plate-forme, notamment la base tarifaire associée à la pondération. La plate-forme distante comprend en outre des premiers moyens de traitement informatique permettant d'établir, au moment des paiements, une

10 connexion informatique entre le serveur de la plate-forme et un serveur de paiement distant de la plate-forme. La plate-forme distante comprend en outre des premiers moyens de transmission pour transmettre au serveur de paiement distant des informations de paiement relatives aux utilisateurs, notamment relatives à un code numérique de carte de paiement, à des

15 coordonnées bancaires, à une procédure de paiement de type SWIFT. Les informations de paiement ont été communiquées, par les utilisateurs à la plate-forme, notamment au moyen des équipements informatiques, via le réseau de communication. Ainsi, le serveur de paiement, ayant reçu les informations de paiement relatives aux utilisateurs, peut faire transférer le

20 montant correspondant au paiement au compte de l'opérateur de la plate-forme, et notifier à ladite plate-forme une information de validation du paiement. Le système comprend des seconds moyens de transmission, notamment associés aux équipements informatiques des utilisateurs, pour transmettre les informations techniques à la plate-forme via le réseau de

25 communication. Le système comprend en outre des moyens d'extraction, notamment des seconds moyens de traitement informatique, pour extraire des informations techniques une formulation déterminée en fonction de critères prédéterminés, notamment en fonction de critères législatifs et/ou réglementaires. Le système comprend en outre des moyens d'assemblage,

notamment des troisièmes moyens de traitement informatique, pour assembler la formulation déterminée et les informations techniques dans un document final se présentant notamment sous la forme d'un fichier informatique au format XSL-FO ou XML/OpenOffice. Le document final ainsi obtenu peut être converti dans d'autres formats informatiques du type PDF ou RTF et imprimé grâce à des moyens d'impression, une imprimante individuelle par exemple. Le système comprend en outre des troisièmes moyens de transmission pour transmettre, notamment via le réseau de communication, le document final à une entité officielle indépendante de la plate-forme. L'entité officielle est destinée à valider, de manière indépendante, la compatibilité du document final avec les critères prédéterminés. L'utilisateur obtient ainsi confirmation de l'entité officielle indépendante de la fonctionnalité du document final généré par la plate-forme.

De préférence, le système est tel que la base de données comprend en outre des fichiers-maquette, mémorisés notamment sous la forme de fichiers au format XML. Les moyens d'assemblage comprennent en outre un module logiciel d'analyse, destiné à détecter, dans lesdits fichiers-maquette, un premier groupe de données sémantiques, notamment des occurrences syntaxiques et/ou linguistiques. Les moyens d'assemblage comprennent en outre un premier module logiciel de traitement sémantique, destiné à générer, à partir du premier groupe de données sémantiques, un flux de données primaires statistiques, notamment des constances d'occurrences et/ou des cycles d'occurrences. Les moyens d'assemblage comprennent en outre un second module logiciel de traitement sémantique, destiné à générer, à partir de la formulation déterminée, un flux de données secondaires en fonction du flux de données primaires statistiques, notamment en adaptant et/ou en modifiant la formulation déterminée en fonction des constances d'occurrences et/ou des cycles d'occurrences.

Ainsi, la fonctionnalité du document final généré par la plate-forme est établie préalablement à sa validation par l'entité officielle indépendante. Il est également possible d'ajouter au second module logiciel de traitement sémantique des fonctionnalités comme le balisage lexical, le balisage sémantique et le balisage analytique. L'étape de codage et de balisage lexical consiste à établir une transcription préliminaire des informations techniques selon des formats numérisés structurés. Le balisage sémantique suit le même principe : l'analyse mot à mot permettra par exemple de mettre en évidence les noms propres et même de régulariser leur orthographe le cas échéant.

De préférence, le système est tel que les moyens d'assemblage, notamment des troisièmes moyens de traitement informatique, comprennent un troisième module logiciel de conversion dynamique de flux, notamment destiné à convertir le document final au format HTML. Les moyens d'assemblage, notamment des troisièmes moyens de traitement informatique, comprennent en outre des moyens de compression pour compresser la formulation déterminée et/ou les informations techniques. Ainsi, la taille du document final généré par la plate-forme est réduite et l'accessibilité distante, notamment via le réseau de communication Internet, du document final généré par la plate-forme est accélérée.

De préférence, le troisième module logiciel de conversion dynamique de flux est associé à une mémoire cache. Ainsi, la publication d'un document final volumineux, notamment excédant 145 Mo, généré par la plate-forme est rendue fluide, y compris en temps réel.

La figure 1 représente, de manière schématique, le système de traitement informatique permettant à des utilisateurs 1, 2 d'un réseau de communication 3, notamment du type Internet, de générer des documents fonctionnels à partir d'informations techniques. Le système de traitement informatique est utilisé dans un système de paiement sécurisé destiné à

permettre aux utilisateurs 1, 2 du réseau de communication 3, notamment du type Internet, d'effectuer des paiements destinés à l'acquisition des documents fonctionnels proposés en vente par une plate-forme distante 4 des utilisateurs 1, 2 du réseau de communication 3. La plate-forme distante 4 du système de paiement sécurisé comprend un serveur 5 qui comporte des premiers moyens de calcul 6 pour calculer, en fonction d'une série de paramètres fixes, une base tarifaire et des seconds moyens de calcul 7 pour calculer, en fonction d'une série de paramètres variables, une fraction variable tarifaire et des moyens de comparaison 8 pour comparer la base tarifaire avec la fraction variable et calculer une pondération en fonction de la comparaison ainsi effectuée. La plate-forme distante 4 comprend en outre une base de données 9 connectée au serveur 5 et des moyens de mémorisation 10 pour mémoriser dans la base de données 9 la série de paramètres fixes et lad série de paramètres variables, ladite base tarifaire associée à ladite pondération. La plate-forme distante 4 est connectée, via le réseau de communication 3 informatique, à des équipements informatiques 11. Les équipements informatiques 11 comportent des moyens de visualisation 11a sur lesquels les utilisateurs 1, 2 peuvent lire des informations diffusées par la plate-forme distante 4, notamment la base tarifaire associée à la pondération. La plate-forme distante 4 comprend en outre des premiers moyens de traitement informatique 12 permettant d'établir, au moment des paiements, une connexion informatique entre le serveur 5 de la plate-forme distante 4 et un serveur de paiement distant 13 de la plate-forme distante 4, des premiers moyens de transmission 14 pour transmettre au serveur de paiement distant 13 des informations de paiement 15 relatives aux utilisateurs 1, 2, notamment relatives à un code numérique de carte de paiement, à des coordonnées bancaires, à une procédure de paiement de type SWIFT. Les informations de paiement 15 ont été communiquées, par les utilisateurs 1, 2 à la plate-forme distante 4,

notamment au moyen des équipements informatiques 11, via le réseau de communication 3. Ainsi, le serveur de paiement, ayant reçu les informations de paiement relatives aux utilisateurs, peut faire transférer le montant correspondant au paiement au compte de l'opérateur de la plate-forme distante, et notifier à la plate-forme distante une information de validation du paiement. Le système comprend des seconds moyens de transmission 16, notamment associés aux équipements informatiques 11 des utilisateurs 1, 2, pour transmettre les informations techniques à la plate-forme distante 4 via le réseau de communication 3, des moyens d'extraction, notamment des seconds moyens de traitement informatique 17, pour extraire des informations techniques une formulation déterminée 23 en fonction de critères prédéterminés, notamment en fonction de critères législatifs et/ou réglementaires. Les seconds moyens de traitement informatique 17 sont paramétrés selon des niveaux de récursivité prédéterminés et selon des tableaux de hachage (ou "tableaux associatifs") qui augmentent dès que nécessaire afin de garantir un bon niveau de performance. Les seconds moyens de traitement informatique 17 peuvent ainsi traiter très rapidement de très grandes quantités de données. Bien qu'optimisé pour le traitement des fichiers textes. Le système comprend en outre des moyens d'assemblage, notamment des troisièmes moyens de traitement informatique 18, pour assembler la formulation déterminée et les informations techniques dans un document final 22 se présentant notamment sous la forme d'un fichier informatique au format XSL-FO ou XML/OpenOffice. Le document final 22 ainsi obtenu peut être converti dans d'autres formats informatiques du type PDF ou RTF et imprimé grâce à des moyens d'impression 11c. Le système comprend en outre des troisièmes moyens de transmission 19 pour transmettre, notamment via le réseau de communication 3, le document final 22 à une entité officielle indépendante 20 de la plate-forme distante 4. L'entité officielle 20 est

destinée à valider, de manière indépendante, la compatibilité du document final 22 avec les critères prédéterminés. De cette manière, l'utilisateur obtient confirmation de l'entité officielle indépendante de la fonctionnalité du document final généré par la plate-forme distante. La base de données 9

5 comprend en outre des fichiers-maquette 21, mémorisés notamment sous la forme de fichiers au format XML. Les moyens d'assemblage comprennent en outre un module logiciel d'analyse 18a, destiné à détecter, dans les fichiers-maquette 21, un premier groupe de données sémantiques, notamment des occurrences syntaxiques et/ou linguistiques, un premier

10 module logiciel de traitement sémantique 18b, destiné à générer, à partir du premier groupe de données sémantiques, un flux de données primaires statistiques, notamment des constances d'occurrences et/ou des cycles d'occurrences, un second module logiciel de traitement sémantique 18c, destiné à générer, à partir de la formulation déterminée 23, un flux de

15 données secondaires en fonction du flux de données primaires statistiques, notamment en adaptant et/ou en modifiant la formulation déterminée 23 en fonction des constances d'occurrences et/ou des cycles d'occurrences. Ainsi, la fonctionnalité du document final généré par la plate-forme distante est établie préalablement à sa validation par l'entité officielle indépendante.

20 Les moyens d'assemblage, notamment des troisièmes moyens de traitement informatique 18, comprennent un troisième module logiciel de conversion dynamique de flux 18d, notamment destiné à convertir le document final 22 au format HTML, des moyens de compression 18e pour compresser la formulation déterminée 23 et/ou les informations techniques. Les moyens

25 de compression 18e peuvent par exemple se fonder sur des techniques de quantification vectorielle permettant la quantification vectorielle "temps réel" et donc la possible émission et réception de trames "temps réel" avec de bas débits. Les moyens de compression 18e peuvent également consister, de manière connue en soi, en une technique de compression

sélective permettant une compression plus ou moins grande selon la taille de la trame. Quelle que soit la technique utilisée, la taille du document final généré par la plate-forme distante est ainsi réduite et l'accessibilité distante, notamment via le réseau de communication Internet, du document final  
5 généré par la plate-forme distante est accélérée. Le troisième module logiciel de conversion dynamique de flux 18d est associé à une mémoire cache 18f, de sorte que la publication d'un document final volumineux, notamment excédant 145 Mo, généré par la plate-forme distante est rendue fluide, y compris en temps réel.

10 On va maintenant décrire une variante de réalisation de l'invention dans laquelle la base des informations techniques est, pour mesurer la puissance du système, constituée par un corpus que l'on suppose relativement volumineux, notamment dépassant 100 000 caractères (hors espaces). Le système selon l'invention permet un traitement préliminaire selon trois  
15 axes : mots, documents et thèmes. Dans cette navigation, dite triangulaire, une liste de mots résulte d'une combinaison booléenne de documents et/ou de thèmes, une liste de documents résulte d'une combinaison booléenne de mots et/ou de thèmes. Les modules logiciels des moyens d'extraction permettent de construire une première indexation de départ. De manière  
20 dynamique, au cours de la navigation, l'utilisateur peut modifier l'indexation : suppression, modification, création, fusion de mots d'index. Les modules logiciels des moyens d'extraction permettent en outre aussi la suppression et la modification d'unités documentaires et même la création de nouvelles unités. Grâce à un accès lexical au vocabulaire d'indexation,  
25 la saisie d'un mot par l'utilisateur engendre la liste des mots d'index du corpus les plus proches au plan lexical. Cette fonctionnalité apporte la nécessaire tolérance aux variantes d'écriture d'un même mot ou aux fautes d'orthographe, qu'elles viennent de l'utilisateur ou du corpus lui-même, cette tolérance devant être appréciée à la fois de manière objective et

subjective en tenant compte des exigences du destinataire final du document fonctionnel. L'expansion sémantique engendre ensuite la liste des documents proches sémantiquement d'un ou plusieurs documents donnés (les documents qui partagent le plus de termes significatifs communs) ainsi que la liste des mots d'index proches sémantiquement d'un ou plusieurs mots donnés (les mots le plus souvent associés dans les documents). Le premier module logiciel de traitement sémantique procède alors à la classification automatique des documents en résumant le corpus en ses composantes thématiques principales. Ces composantes évoluent au fil des modifications opérées par l'utilisateur sur l'indexation du corpus documentaire. A chacun des thèmes sont associés une liste de documents, ordonnés par importance décroissante et une liste de mots clés également ordonnés. Ces listes dont l'homogénéité signe la qualité de la progression sémantique sont mémorisées dans la base de données de la plate-forme et fournissent la matière nécessaire à la compréhension fine des regroupements effectués. Les listes des mots caractéristiques des thèmes identifiés sont analysées par le second module logiciel de traitement sémantique selon une méthode d'analyse en composantes principales pour être présentées dans un plan factoriel. Les proximités géométriques entre les thèmes sont renforcées et nuancées par le dessin des liaisons entre les thèmes proches. A tout moment du traitement, l'utilisateur peut accéder au document en cours d'élaboration, visualisable dans une version provisoire à la mise en page conviviale. Les retours en arrière sont rendus possibles en passant d'une vue sauvegardée à l'autre de façon à faciliter le cheminement vers la vision de la base la mieux adaptée à la problématique de chaque utilisateur. La duplication de vue, le zoom sur un thème et la comparaison de cartes sémantiques sont assurés. La collecte de termes au fur et à mesure des navigations est rendue possible grâce à un panier virtuel, réceptacle de mots autonome et transversal à l'ensemble des vues. Le panier est

exportable vers d'autres applications. Le troisième module logiciel de conversion dynamique de flux permet enfin de publier au format HTML, notamment sur un site Internet tout ou partie des vues sur un corpus textuel. Celles-ci peuvent alors être consultées via Internet, notamment grâce à un applet Java autorisant une navigation triangulaire couplée à une visualisation de cartes sémantiques. On peut mentionner de nombreux types de réalisations potentielles de documents présentant des degrés et des natures de fonctionnalité variés : rapports de synthèses, cahiers des charges multi-domaines soumis à validation, formulaires et bordereaux dans le domaine pharmaco-médical, demandes de brevets d'inventions, relevés d'audits à usage juridico-administratif, rapports d'expertises. L'architecture du système selon la présente invention présente par ailleurs l'avantage d'offrir potentiellement, outre la stricte création de documents fonctionnels, la génération d'un langage distribué permettant une méga programmation c'est-à-dire la possibilité de coordonner et d'harmoniser de multiples calculs sur de multiples machines. Pour cela, il est nécessaire d'étudier en particulier, à des niveaux très divers, le domaine des mémoires partagées réparties, des glaneurs de cellules, des couplages avec des bibliothèques existantes. La réalisation d'applications coopératives passe par la disponibilité d'une mémoire partagée répartie. Un langage multi-tâches et distribué est extrêmement sensible au type de cohérence qu'apporte ou qu'impose la mémoire sous-jacente. Pour définir une sémantique compréhensible d'un tel langage, il est nécessaire de spécifier très précisément les propriétés d'une telle mémoire et des protocoles d'accès. Cette sémantique peut être développée dans plusieurs formalismes (dénotationnel, calcul de processus). Une mémoire répartie impose la présence d'un module récupérateur. La méga programmation requiert l'interopérabilité entre langages et/ou bibliothèques, c'est-à-dire la faculté de mêler des données et des programmes obéissant à des conventions de

représentation et de manipulation différentes. Cela suppose la mise en place de gestionnaires de mémoire, contrôlant l'allocation, l'accès et le recyclage des données, qui soient neutres, autrement dit, qui s'adaptent aux besoins spécifiques des utilisateurs. Les questions de fiabilité et de sûreté sont certainement les points les plus critiques à traiter dans le développement d'un tel module logiciel complémentaire, en particulier lorsqu'il s'agit d'un système construit par coopération d'applications. On décompose généralement le problème en deux parties : tout d'abord, il faut construire une spécification formelle à partir du cahier des charges en s'assurant qu'elle répond bien aux exigences de ce cahier. Il faut ensuite montrer que l'implantation proposée est correcte vis-à-vis de cette spécification. Toute solution passe donc d'abord par la conception de langages de spécification et de programmation, qui soient sémantiquement fondés, condition minimale pour pouvoir exprimer et prouver des propriétés. Ensuite, un tel module ne saurait se dispenser de l'adjonction de systèmes d'aide à la vérification et/ou à la preuve. Il existe actuellement un certain nombre de systèmes opérationnels d'aide à la preuve, fondés sur des logiques et des méthodologies différentes. Si les fondements logiques de ces systèmes sont assez bien compris, les outils qui en dérivent sont encore très rudimentaires et la méthodologie de leur emploi encore sommaire. Faire une preuve aidée suppose une formalisation des flux de données et donc une certaine maîtrise des fondements logiques du système. Son utilisation nécessite par ailleurs une bonne compréhension de sa structure interne et plus généralement, la construction de la preuve s'apparente à une construction de programme et demande des compétences avancées en programmation. Le domaine de la preuve aidée par les moyens de traitement informatique modernes peut notamment être intégré à différents systèmes opérationnels existants d'aide à la preuve (B, Coq, PVS), au travers applications de nature différente conjointement dans

plusieurs systèmes, de manière à comparer leur faisabilité et à élaborer une méthodologie de développement pour chaque type d'application.

Un autre avantage du système selon l'invention est de pouvoir intégrer de manière transversale des modules fonctionnels complémentaires autorisant une analyse semi-automatique des informations techniques brutes à partir desquelles le document final doit être généré. Le système selon une autre variante de réalisation de l'invention comprend des modules complémentaires, notamment des modules de désambiguïsation grammaticale (déterminer, par exemple, si "ferme" est un nom, un adjectif, un adverbe ou une forme verbale), des modules d'analyse syntaxique et logique de la phrase (découper les propositions, constituer les groupes grammaticaux, déterminer les fonctions grammaticales comme le sujet, l'attribut, l'apposition), des modules de désambiguïsation sémantique (pour les mots ayant plusieurs sens, déterminer le sens de chacun de ces mots dans le contexte), des modules de repérage des expressions et des entités nommées (regroupement des groupes nominaux propres comme "Société nationale d'exploitation industrielle des tabacs et allumettes"), des modules de résolution des anaphores (retrouver le référent des pronoms comme "il", "en" ou "ce dernier", et d'autres termes comme "le premier") des modules d'analyse de thèmes (déterminer le sujet, les thèmes du texte, du paragraphe, de la phrase). En outre, des modules de détection des têtes de dérivation, de repérage des métaphores, de correction automatique des fautes d'orthographe sont également inter-opérables avec le système selon l'invention. L'adjonction de ces modules préserve la nécessaire segmentation fonctionnelle entre recherche et extraction d'informations, les moyens d'extraction d'informations proposant des membres de phrases répondant directement aux requêtes des utilisateurs.

### Revendications

1. Système de traitement informatique permettant à des utilisateurs (1, 2) d'un réseau de communication (3), notamment du type Internet, de générer des documents fonctionnels à partir d'informations techniques ;
- 5 ledit système de traitement informatique étant utilisé dans un système de paiement sécurisé destiné à permettre auxdits utilisateurs (1, 2) dudit réseau de communication (3), notamment du type Internet, d'effectuer des paiements destinés à l'acquisition desdits documents fonctionnels proposés en vente par une plate-forme distante (4) desdits utilisateurs (1, 2)
- 10 dudit réseau de communication (3) ;
- ladite plate-forme distante (4) dudit système de paiement sécurisé comprenant un serveur (5) ; ledit serveur (5) comportant :
- des premiers moyens de calcul (6) pour calculer, en fonction d'une série de paramètres fixes, une base tarifaire,
  - 15 - des seconds moyens de calcul (7) pour calculer, en fonction d'une série de paramètres variables, une fraction variable tarifaire,
  - des moyens de comparaison (8) pour :
    - comparer ladite base tarifaire avec ladite fraction variable,
    - calculer une pondération en fonction de ladite comparaison ainsi
- 20 effectuée,
- ladite plate-forme distante (4) comprenant en outre :
- une base de données (9) connectée audit serveur (5),
  - des moyens de mémorisation (10) pour mémoriser dans ladite base de données (9) :
- 25
  - ladite série de paramètres fixes et ladite série de paramètres variables,
  - ladite base tarifaire associée à ladite pondération ;
- ladite plate-forme distante (4) étant connectée, via ledit réseau de communication (3) informatique, à des équipements informatiques (11) ;

lesdits équipements informatiques (11) comportant des moyens de visualisation (11a) sur lesquels lesdits utilisateurs (1, 2) peuvent lire des informations diffusées par ladite plate-forme distante (4), notamment ladite base tarifaire associée à ladite pondération ;

5 ladite plate-forme distante (4) comprenant en outre :

- des premiers moyens de traitement informatique (12) permettant d'établir, au moment desdits paiements, une connexion informatique entre ledit serveur (5) de ladite plate-forme distante (4) et un serveur de paiement distant (13) de ladite plate-forme distante (4),

10 - des premiers moyens de transmission (14) pour transmettre audit serveur de paiement distant (13) des informations de paiement (15) relatives auxdits utilisateurs (1, 2), notamment relatives à un code numérique de carte de paiement, à des coordonnées bancaires, à une procédure de paiement de type SWIFT ;

15 lesdites informations de paiement (15) ayant été communiquées, par lesdits utilisateurs (1, 2) à ladite plate-forme distante (4), notamment au moyen desdits équipements informatiques (11), via ledit réseau de communication (3) ;

de sorte que le serveur de paiement, ayant reçu les informations de paiement relatives aux utilisateurs, peut :

20 • faire transférer le montant correspondant au paiement au compte de l'opérateur de la plate-forme distante, et

• notifier à la plate-forme distante une information de validation du paiement ;

25 ledit système comprenant :

- des seconds moyens de transmission (16), notamment associés auxdits équipements informatiques (11) desdits utilisateurs (1, 2), pour transmettre lesdites informations techniques à ladite plate-forme distante (4) via ledit réseau de communication (3),

- des moyens d'extraction, notamment des seconds moyens de traitement informatique (17), pour extraire desdites informations techniques une formulation déterminée (23) en fonction de critères prédéterminés, notamment en fonction de critères législatifs et/ou réglementaires,

5 - des moyens d'assemblage, notamment des troisièmes moyens de traitement informatique (18), pour assembler ladite formulation déterminée et lesdites informations techniques dans un document final (22) se présentant notamment sous la forme d'un fichier informatique au format XSL-FO ou XML/OpenOffice ; ledit document final (22) ainsi obtenu  
10 pouvant être converti dans d'autres formats informatiques du type PDF ou RTF et imprimés grâce à des moyens d'impression (11c) ;

ledit système comprenant en outre des troisièmes moyens de transmission (19) pour transmettre, notamment via ledit réseau de communication (3), ledit document final (22) à une entité officielle indépendante (20) de ladite  
15 plate-forme distante (4) ; ladite entité officielle (20) étant destinée à valider, de manière indépendante, la compatibilité dudit document final (22) avec lesdits critères prédéterminés, de sorte que l'utilisateur obtient confirmation de l'entité officielle indépendante de la fonctionnalité du document final généré par la plate-forme distante.

20 **2.** Système selon la revendication 1 ; ledit système étant tel que ladite base de données (9) comprend en outre des fichiers-maquette (21), mémorisés notamment sous la forme de fichiers au format XML ; lesdits moyens d'assemblage comprenant en outre :

25 - un module logiciel d'analyse (18a), destiné à détecter, dans lesdits fichiers-maquette (21), un premier groupe de données sémantiques, notamment des occurrences syntaxiques et/ou linguistiques,

- un premier module logiciel de traitement sémantique (18b), destiné à générer, à partir dudit premier groupe de données sémantiques, un flux de

données primaires statistiques, notamment des constances d'occurrences et/ou des cycles d'occurrences,

5 - un second module logiciel de traitement sémantique (18c), destiné à générer, à partir de ladite formulation déterminée (23), un flux de données secondaires en fonction dudit flux de données primaires statistiques, notamment en adaptant et/ou en modifiant ladite formulation déterminée (23) en fonction desdites constances d'occurrences et/ou desdits cycles d'occurrences, de sorte que la fonctionnalité du document final généré par la plate-forme distante est établie préalablement à sa validation par l'entité  
10 officielle indépendante.

3. Système selon l'une quelconque des revendications 1 ou 2 ; ledit système étant tel que lesdits moyens d'assemblage, notamment des troisièmes moyens de traitement informatique (18), comprennent :

15 - un troisième module logiciel de conversion dynamique de flux (18d), notamment destiné à convertir ledit document final (22) au format HTML,

20 - des moyens de compression (18e) pour compresser ladite formulation déterminée (23) et/ou lesdites informations techniques, de sorte que la taille du document final généré par la plate-forme distante est réduite, de sorte que l'accessibilité distante, notamment via le réseau de communication Internet, du document final généré par la plate-forme distante est accélérée.

4. Système selon la revendication 3 ; ledit système étant tel que ledit troisième module logiciel de conversion dynamique de flux (18d) est associé à une mémoire cache (18f), de sorte que la publication d'un  
25 document final volumineux, notamment excédant 145 Mo, généré par la plate-forme distante est rendue fluide, y compris en temps réel.

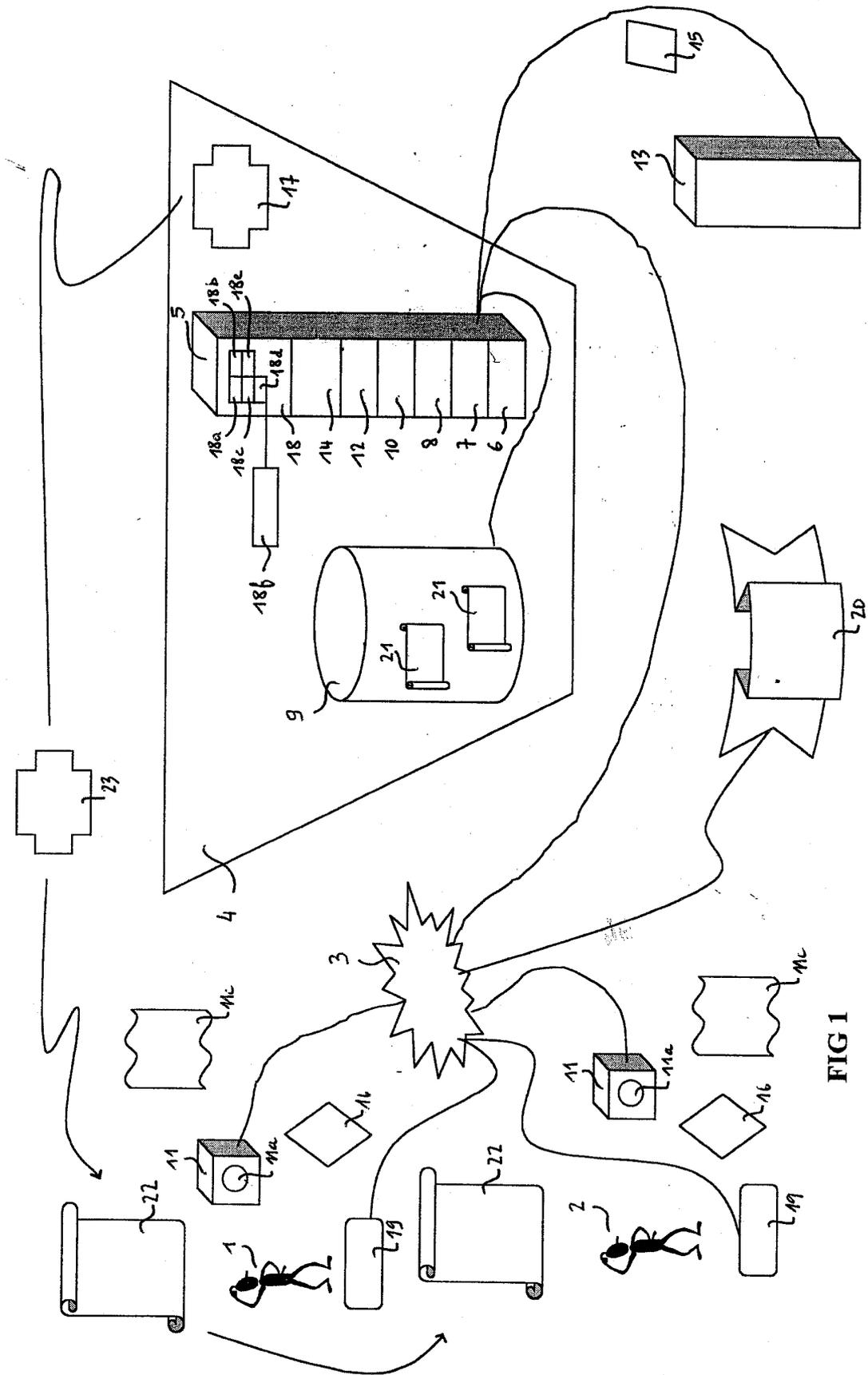


FIG 1



**ANNEXE AU RAPPORT DE RECHERCHE PRÉLIMINAIRE  
RELATIF A LA DEMANDE DE BREVET FRANÇAIS NO. FR 0303222 FA 633449**

La présente annexe indique les membres de la famille de brevets relatifs aux documents brevets cités dans le rapport de recherche préliminaire visé ci-dessus.

Les dits membres sont contenus au fichier informatique de l'Office européen des brevets à la date du 21-01-2004

Les renseignements fournis sont donnés à titre indicatif et n'engagent pas la responsabilité de l'Office européen des brevets, ni de l'Administration française

Document brevet cité au rapport de recherche	Date de publication	Membre(s) de la famille de brevet(s)	Date de publication
WO 9931610 A	24-06-1999	AU 753258 B2	10-10-2002
		AU 1570399 A	05-07-1999
		CA 2313388 A1	24-06-1999
		CN 1282432 T	31-01-2001
		DE 69811947 D1	10-04-2003
		DE 69811947 T2	11-12-2003
		EP 1040437 A1	04-10-2000
		WO 9931610 A1	24-06-1999
		JP 2002509300 T	26-03-2002
		NZ 505057 A	25-10-2002
		US 2002178070 A1	28-11-2002
EP 1085478 A	21-03-2001	JP 11272565 A	08-10-1999
		JP 11282867 A	15-10-1999
		EP 1085478 A1	21-03-2001
		CN 1295699 T	16-05-2001
		WO 9949430 A1	30-09-1999
US 2002107853 A1	08-08-2002	AUCUN	
WO 01063485 A	30-08-2001	AU 4321801 A	03-09-2001
		WO 0163485 A2	30-08-2001
US 2001056497 A1	27-12-2001	CN 1330326 A	09-01-2002
EP 1187079 A	13-03-2002	FR 2813734 A1	08-03-2002
		EP 1187079 A1	13-03-2002
FR 2808907 A	16-11-2001	FR 2808907 A1	16-11-2001
		AU 6040401 A	20-11-2001
		WO 0186496 A1	15-11-2001