

Le travail de groupe sur l'Internet : une application dans le domaine de la veille technologique

André Le Meur
Laboratoire CRAIE
Université de Rennes 2
5, avenue Gaston Berger, 35043 Rennes Cedex
andre.lemeur@uhb.fr

octobre 1995

Résumé

Les services qui ont fait le succès du réseau Internet portent essentiellement sur l'**échange** d'information entre individus (le courrier électronique) ou sur leur **diffusion** ("news", "liste", "gopher", "W3" etc.). La recherche présentée ici porte sur la **définition** d'un autre type de service : le **travail collectif** dans lequel un les membres d'un groupe de contributeurs géographiquement dispersés alimentent un **réservoir commun** d'informations, y puisent, ajoutent leurs commentaires etc.

Inspiré des "réservoirs de notices" des réseaux de catalogage partagé, le réseau BALNEO de collecte des néologismes de la langue française est développé au laboratoire CRAIE de l'Université de Rennes 2 pour le compte du RINT (Réseau International de Néologie et de Terminologie).

Après un examen des attentes des commanditaires et des usagers, la description de la première version de ce produit permettra de voir l'utilisation du **courrier électronique** (des Bâtes Aux Lettres pour la NEOlogie) pour exploiter une **base de données centrale** et les problèmes rencontrés : la faible interactivité et les difficultés d'authentification. On verra ensuite à partir de la version en cours de développement l'apport de W3 pour résoudre ces problèmes.

1 - le cahier des charges :

Le but du réseau BALNEO est de **collecter rapidement** et de **diffuser largement** au moyen d'un réseau informatique des attestations de néologismes lus ou entendus par des observateurs francophones. Par **néologisme**, il faut entendre un terme (un mot ou une expression) appartenant en général à un langage de spécialité et qui est soit un nouveau terme, soit un terme ancien qui prend un nouveau sens dans la langue. Par **attestation**, il faut entendre la description des circonstances de cette apparition (le texte, la phrase, la date, l'auteur etc.). Voici quelques unes des caractéristiques attendues d'un tel réseau :

La demande institutionnelle

- le réseau informatique doit se calquer sur la **structure hiérarchique** du réseau commanditaire qui est composé de **modules** nationaux qui travaillent avec des **organismes** dans lesquels des **rédacteurs** sont chargés de la production des "attestations". Chaque instance hiérarchique peut valider (et invalider) une instance de niveau inférieur et s'en porte garant.
- la gestion du réseau doit être automatisée au maximum : après son inscription (par une instance supérieure) sur le réseau, tout membre ajoute, consulte et commente librement les attestations sans intervention d'une tierce personne (il n'y a ni censeur ni "modérateur").

- la description d'un néologisme et des conditions de son apparition ayant fait l'objet préalable au sein du RINT, le réseau doit collecter et faire circuler des "fiches" structurées conformes à un modèle préétabli défini par une DTD SGML. L'information échangée sur le réseau est purement textuelle (pas de son, de figure, de tableau.... Le jeu de caractères utilisé est limité à l'ISO-latin.
- le réseau doit garantir que toute information diffusée en mentionne le rédacteur, le titulaire des droits de réutilisation, et informe sur les conditions de réexploitation et de cession.
- seul le rédacteur d'une information peut la modifier mais une instance supérieure peut se substituer à lui en cas de d'abus ou de défaillance.
- tout membre du réseau peut **annoter** une information produite par un autre membre. Les annotations sont jointes à l'information originale lors de sa diffusion. Elles ne peuvent être modifiées ou supprimées que par leur auteur.

Les usagers

- les membres du réseau sont très dispersés (au moins en Europe, en Amérique et en Afrique) mais il sont supposés avoir un accès à Internet à partir de micro-ordinateurs sous Windows.
- le travail de saisie et de mise à jour de l'information est long, il doit s'effectuer "hors ligne" avec exportation périodique des données vers la base centrale.
- la consultation de cette base et la rédaction d'annotations doivent pouvoir se faire "hors ligne" et "en ligne".
- les rédacteurs sont essentiellement des utilisateurs de traitement de texte. Les outils locaux de saisie et les procédures d'import/export via Internet doivent donc être simples et conformes aux règles ergonomiques de Windows qu'ils connaissent déjà.

2 - Principes de la première version (échanges en différé par messagerie)

La production de l'information

La première version de BALNEO utilise uniquement une "base éditoriale" pour la saisie des néologismes et la génération des messages administratifs. Cette base développée à partir de ACCESS de MicroSoft permet la saisie des néologismes, leur recherche, leur mise à jour... et la formulation de requêtes.

L'envoi de l'information

Une fonction d'exportation exporte la base de données ou la requête dans un fichier ou dans le presse-papiers de Windows. Le résultat de cette exportation est transmis par un logiciel de messagerie (Mail, Eudora...) à une boîte aux lettres située sur un serveur central.

Le traitement des messages

Sur le serveur central un programme capte le message (selon le principe des "serveurs de liste" qui a inspiré cette première version). Il identifie l'émetteur (à partir du champ FROM) et interprète le contenu du message. Si c'est une commande d'ajout, il enregistre les données jointes au message dans une base de données relationnelles. S'il s'agit d'une requête, il interroge la base centrale et renvoie les réponses dans la boîte aux lettres de l'émetteur du message. Celui-ci lit son courrier, extrait la réponse dans un fichier ou le presse-papiers puis l'importe dans sa base éditoriale locale ou il peut lire, annoter, éditer...

Avantages et inconvénients de cette méthode

Le procédé est simple à mettre en oeuvre. Il se prête bien à la production des données (travail long sur la base locale puis envoi périodique en masse) mais il est lent et lourd pour les interrogations ponctuelles de la base centrale (pour vérifier par exemple si un néologisme a déjà été enregistré). Le défaut majeur de la messagerie réside dans le procédé d'identification de l'émetteur : les logiciels de messagerie sous Windows ne contrôlent pas le contenu du champ FROM, cette information ne peut donc suffire à authentifier l'émetteur.

3 - Les évolutions : l'accès par W3

Peu après la conception de cette première version, le standard HTML 2+ qui est à la base des services W3 a apporté un nouvel élément pour la conception de services interactifs : **les formulaires**. Grâce à eux, l'utilisateur d'un client W3 peut transmettre des informations à un serveur W3 où elles sont immédiatement traitées par un programme. Une nouvelle version de BALNEO est donc en cours de développement. Le système de **messagerie** est conservé pour les **échanges massifs** de données (l'authentification est assurée par la technique du message de confirmation) mais les fonctions d'**administration**, de **consultation rapide** et d'**annotation** passent par des formulaires W3. Voici quelques unes de ces fonctions□:

L'authentification des utilisateurs

Tout utilisateur reçoit lors de son inscription sur le réseau un mot de passe. Par la suite quand il se connecte au moyen d'un client W3, il fournit ce code confidentiel. Une "clef" est calculée et transmise (par l'intermédiaire de champs cachés) au client W3 qui la renvoie au serveur lors de chaque transaction. Cette clef est imprévisible, indéchiffrable et a une durée de vie limitée. On peut penser que la fiabilité de ce procédé est suffisante dans ce contexte d'utilisation.

Administration, consultation et saisie rapides, annotations...

L'utilisateur authentifié peut procéder à toutes les opérations administratives selon le statut et les privilèges qu'il a dans le réseau.

Il peut formuler une requête sur la base centrale et demander à voir la réponse immédiatement ou à la recevoir par messagerie. Quand il lit une fiche, il peut taper une annotation qui sera immédiatement ajoutée à cette fiche.

Un autre formulaire permet une saisie rapide et en ligne d'une attestation mais les limites de l'éditeur HTML rendent l'opération difficile pour certains types de champs complexes et répétitifs.

Quelques autres fonctions

Les éditeurs

La première version de BALNEO ne disposait que de la base éditoriale pour produire de l'information. Les utilisateurs ont souhaité disposer d'outils de saisie plus proches de leur logiciel favori : le traitement de texte. Une étude est en cours pour évaluer les stratégies possibles de saisie contrôlée (utilisation de boîtes de dialogue pilotées par macro-commande, marquage de la structure par les styles etc.).

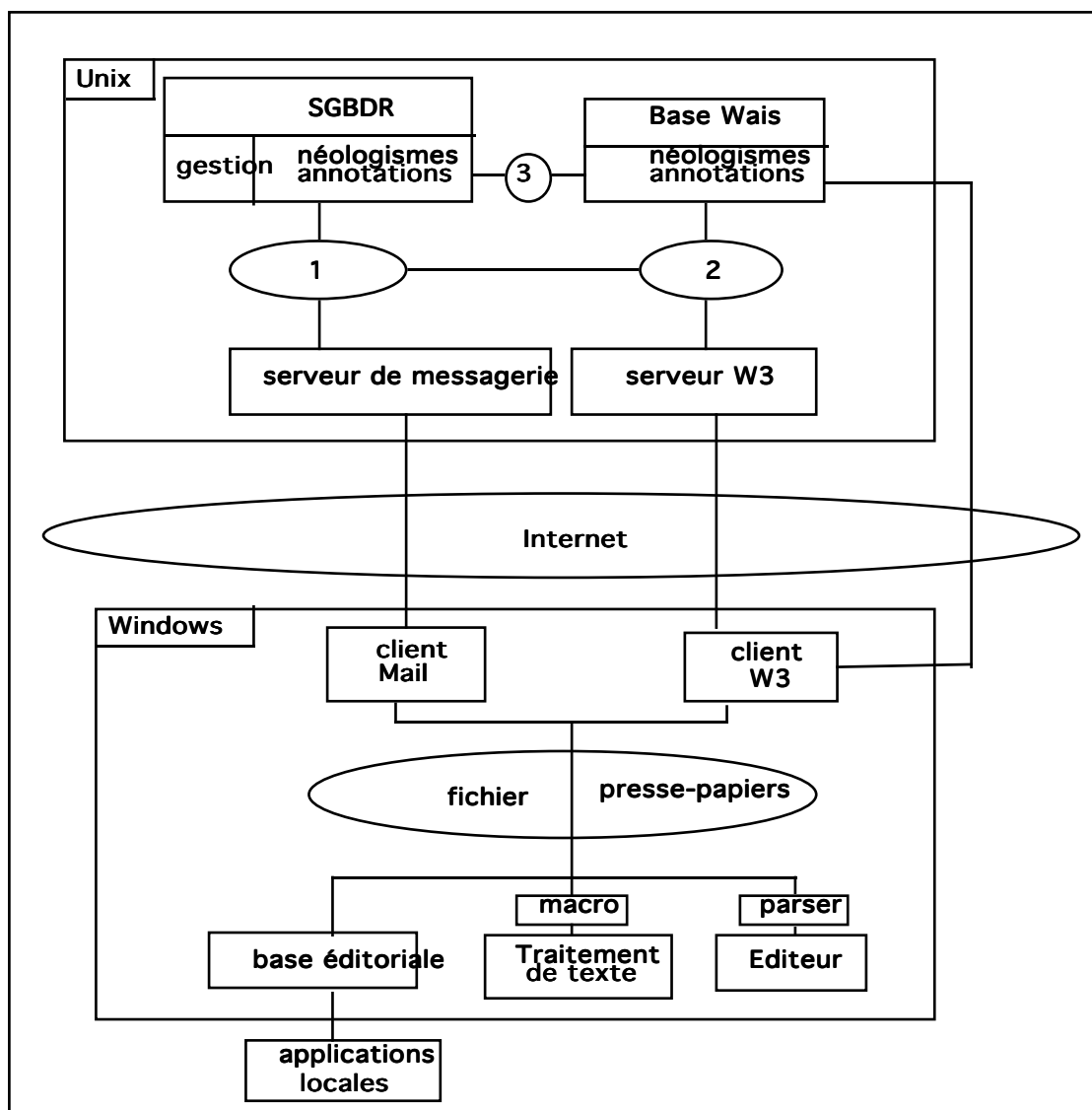
Les applications locales

Les données saisies ou importées dans la base éditoriale peuvent être exportées en format SGML, RTF ou ISO 2709 pour une réutilisation par une base de données documentaire, un logiciel de PAO, une base diffusable sur CDROM etc.

L'utilisation de WAIS

Sur le site central, la gestion des données administratives et néologiques est assurée par un SGBD relationnel (Oracle 7 en l'occurrence). Ce type de base est nécessaire pour la gestion courante (ajouts, mises à jour, suppressions) mais il présente des limites dans la consultation des néologismes par exemple la recherche des mots qui composent un terme). Les néologismes sont donc transférés périodiquement dans une base WAIS. Un formulaire de recherche permet de formuler des équations de recherche plus fines à partir du client W3.

BALNEO : vue d'ensemble



Sur le site central trois applications ont été programmées :

- (1) reçoit, authentifie et traite les messages provenant de la boîte aux lettres ou du programme (2). Il met à jour la base de données relationnelle, renvoie un message à la boîte aux lettres du demandeur ou rend un message à (2)
- (2) authentifie un demandeur, transmet ses demandes à (1), rend les message de retour au client W3.
- (3) convertit la base de néologismes au format WAIS et produit une base interrogeable par (2)

