

MULTILINGUAL INTERNET

Multilingual Distributed Referential
System

networks result from common adherence to accepted standards, parameters, addresses, and names. This is why its referential system in turn builds a network.

INTERNET MULTILINGUE

Système de Référentiels Distribué
Multilingue

les réseaux résultent d'une adhésion commune à des standards, paramètres, adresses et noms connus. C'est pourquoi son système référentiel construit un réseau.

INTRODUCTION GENERALE

En coopération avec nombre de partenaires, INTLNET s'est engagé dans un projet de collection, gestion et dissémination de l'information nécessaire au déploiement d'un **Internet Multilingue** par l'usage : le **SRDM/MDRS** (système de référentiels distribué multilingue/multilingual distributed referential system).

Ce projet se fonde sur une analyse de l'écosystème relationnel numérique résumée ici.

- il y a trois niveaux techniques du réseau numérique :
 - **interconnectivité** (les lignes et commutateurs : infrastructure télécoms),
 - **interopérabilité** (superstructure internet, MPEG, portables téléphone),
 - **interintelligibilité** (métastructure des espaces relationnels entre personnes).

INTLNET utilise une modélisation étendue des réseaux numériques qui développe ces trois niveaux principaux, englobe les couches du modèle OSI, et documente l'intelligence des usages partagés en réseau (notion de noogitel/brainware).

- il y a quatre niveaux linguistiques :
 - **universalisation** (indépendance de la langue, en général numérique),
 - **linguisation** (technologie pour une langue donnée [anglais pour Internet, français pour Minitel]),
 - **globalisation** (extension de la linguisation pour réduire les barrières avec d'autres langues par l'internationalisation [Unicode] du médium et la localisation des terminaux),
 - **multilinguisation** (globalisation de toutes les langues). Le premier besoin est une identification des langues qui soit claire à tous leurs locuteurs, hommes et machines, et donc à leurs espaces d'échanges.

- il y a cinq niveaux de communication humaine (et de gestion de l'autorité - gouvernance) qui se mêlent à des degrés divers selon les densités technico-sociales.
 - **écoute** : j'obéis
 - **soliloque** : je décide pour moi
 - **monologue** : je décide pour les autres
 - **dialogue** : je décide avec les autres
 - **polylogue** : je décide pour moi en tenant compte de ce que les autres font et disent

La convergence technique et multilingue actuelle est une généralisation de la capacité de polylogue globale dont nous avons le besoin, un début d'expérience, mais pas encore les moyens appropriés (exemple : le spam qui provient d'une incapacité technique à contrôler l'ubiquité du système) ni la mécanique d'intergouvernance nécessaire, étant toujours régi par une culture de gouvernance unique.

- le problème rencontré par l'Internet est son unilatéralisme technique, linguistique et "gouvernantal" hérité de son origine académique américaine qui le conduit à se développer comme un monolithe qui - faute d'interopérabilité imaginée ou voulue - doit remplacer les autres technologies en restant sous la coupe du IANA et en étant limité par sa centralisation Deux schémas sont possibles face à cette rigidité et en raison des oppositions qu'elle soulève:
 - la **fragmentation** technique (NATs, instabilité du DNS, etc.) plus ou moins corrigée par l'autoritarisme contesté de l'ICANN,
 - ou le **découplage** du développement des différents niveaux techniques, linguistiques et relationnels permettant une progression multilatérale concertée.
- la multilinguisation que nous envisageons se situe donc **au niveau métastructurel et de l'usage**, indépendante des technologies réseau (le développement est commun, assurant une continuité totale, pour l'Internet, les médias, le téléenseignement, l'e-commerce, le web, les mobiles, etc.).

Il s'agit donc d'identifier, documenter et e-supporter les espaces relationnels linguistiques (mailing lists, foras, lectorats, zones de chalandise, espaces culturels, bassins économiques, zones académiques, résidentielles, professionnelles, etc.).

Pour parvenir à cette architecture des **réseaux du réseau des réseaux**, il faut une technologie d'interface intelligent entre le niveau application des diverses technologies de réseau (OSI layer 7) et les applications utilisateur (serveur). Ce niveau "interapplicatif" devra compatible avec chaque technologie de réseau et potentiellement chaque sémantique, les rendant transparent au contenu. C'est qu'en partie explore le middleware actuel et que l'IAB appelle l'"edge".

Pour la mise en place concrète de l'Internet Multilingue par l'usage, INTLNET s'est engagé dans l'organisation d'un **MLTF (Multilingual Technical Forum)** pour viser quatre axes prioritaires :

- la "cartographie virtuelle" des **espaces relationnels géo/socio-linguistiques** où s'inscrivent les échanges humains et homme/machine. Le but est de parvenir à un registre "langroot" donnant accès de façon structurée, repérée, croisée et dénommée au plus grand nombre de paramètres. Ces paramètres seront actualisés à partir de leur

description par le plus grand nombre de sources (SIL, Linguasphere, ISO, Unicode, IETF, WikiPedia, Francophonie, etc.).

Ce projet utilise le travail sources ouvertes de David Dalby (<http://langtag.com>). Chaque espace disposera d'un site, d'un groupe de travail et d'une base documentaire.

- **le chemin d'e-capacitation linguistique**

pour bien comprendre l'ensemble des besoins devant être supportés et recenser les moyens déjà à disposition à travers les standards ISO, les différents SSDOs, les conditions de leur interoperabilité (ex. IETF : langtags et IDNs), l'évaluation des possibilité techniques et des besoins industriels pour supporter l'e-emergence d'une langue.

C'est le travail engagé à partir de contributions comme celles de Marcel Diki-Kidiri.

- une **réflexion technique** destinée à supporter la mise à disposition des utilisateurs des solutions logicielles et des bases d'information nécessaire à une perception multilingue (la langue de chacun étant SA langue pivot dans le monde numérique).

- **jardins ontologiques :**

forets d'ontologies interliées conditionnellement (intelligence) sur la base des normes ISO, JTC1/SC32/WG2 et une réflexion adaptées au multilinguisme et à la nature distribuée des relations numérisées. Ceci doit permettre d'enregistrer, obtenir, et utiliser toute ontologie informatisée compatible.

Cette partie repose probablement sur de nouveau concepts architecturaux basé sur un nouveau paradigme de compréhension de l'information lorsqu'elle est distribuée, librement gérée par chacun, et multilingue. (Groupe de réflexion France)

- **diffusion et mise à disposition universelle des données**

Les données de ces ontologies documentent le système relationnel commun. Il s'agit d'un bien commun universel dont le DNS fait partie : il ne faudrait en recopier l'inutile et même préjudiciable centralisation actuelle.

Il faut donc : un modèle de compilation ouverte, des outils de gestion, des centres de référence, une diffusion distribuée non propriétaire, un modèle économique pour la diffusion compatible d'informations propriétaires, et des contributions publiques pour les informations de première nécessité (santé, sécurité, alarmes) et d'intérêt général.

Le premier bloc de construction est le "langroot" : la compilation croisée et ouverte des registres linguistiques permettant l'organisation linguistique du projet lui-même (Réunion envisagée à ce sujet début Octobre pour les parties intéressées).

- **organisation pratique de l'e-capacitation individuelle de chacun**

l'autonomie de chacun dans :

- le choix de sa langue
 - la sélection de ses référents (bases d'information "locales" communes à un

- espace linguistique [ex. dictionnaire de l'Académie Française])
- la définition de **ses contextes** personnels et relationnels (variations propres à chacun des référents qu'il utilise) et des autorisations à leur accès
- la protection de ses **manifestes** (descriptif des objets)

demande la spécification et le développement d'outils et de services. Ils devront permettre des interactions transparentes aux technologies utilisées et aux autres méthodes d'exploitation de l'écosystème numérique. De cette façon ils permettront une évolution séparée de chaque système et la stabilité de leur interopérabilité.

(Prototypage en cours).

- la documentation du **premier niveau du nommage multilingue**

Il existe différents services de nommage multilingue

- noms de domaine multilingues (ex. Chine)
- mots clés (Corée, Turquie)
- alias (mots clés privés)
- partie gauche des adresses e-mail
- adresses numériques

Selon une mise à jour quotidienne, INTLNET décrit depuis plusieurs années déjà la zone primaire du DNS telle qu'elle est réellement exploitée. Elle a été étendue à l'Internet Multilingue. Elle comprend en particulier :

- la production d'un **fichier racine de référence** incluant les registres et les serveurs de nommage des domaines linguistiques,

<http://intl.net.org/intlfile.txt>

<http://intl.net.org/intlroot.txt>

- et les **fichiers "bouture"** (*stubs*) nécessaires à leur utilisation en complément du fichier racine de l'ICANN.

L'utilisation de la fonction « punycode » permet une conversion ISO 10646/Hexatridécimal (chiffres de 0 à Z et les signes « . - + »). Cette fonction est en général utilisée pour le support des suffixes de zone primaire et des adresses e-mail linguistiques.