



Serveur Vocal d'Accès à la recherche d'itinéraires Multimodaux en Ile de France

Rapport d'avancement Octobre 2005

Contexte : L'Information Voyageurs figure aujourd'hui parmi les axes prioritaires des acteurs du transport. Mieux informer c'est permettre aux usagers d'être « acteurs » de leurs déplacements privés ou professionnels sur tous les réseaux de transports publics. Pour répondre à cet enjeu important de l'information multimodale, les principales entreprises du transport proposent d'ores et déjà différents médias tel Internet, le Minitel ou encore l'i-mode pour accéder à des services de recherche d'itinéraires urbains et interurbains en Ile de France (*citefutee.com, transport-idf.fr, idf.sncf.fr*).

Néanmoins, ces services d'accès à l'information restent encore trop confidentiels pour répondre à la demande d'un très large public et ne satisfont pas l'exigence de rapidité d'accès, en tous lieux et à tout moment (« situation de mobilité »). Grâce aux technologies qui lui sont associées, la téléphonie, fixe ou mobile, peut apporter des solutions. En effet, les Serveurs Vocaux Interactifs (SVI) nouvelle génération, développés par Vecsys, associent les technologies de la reconnaissance de la parole, de la compréhension du langage parlé et de la synthèse vocale pour offrir à l'utilisateur un accès rapide et intuitif puisque celui-ci communique avec le système par la parole. Ces serveurs sont appelés à juste titre « Serveurs Vocaux Interactifs en Langage Naturel ».

C'est dans ce contexte que la RATP, la SNCF, le LIMSI-CNRS et la société VECSYS ont noué un partenariat pour mener à bien le projet SATIM qui a pour objet la réalisation du prototype d'un Serveur Vocal Interactif (SVI) en langage naturel, qui donne accès par téléphone, à la recherche d'itinéraires multimodaux (réseau ferré RATP et SNCF, tram et bus) pour tout trajet immédiat ou différé en Ile de France.

Objectifs : Plus en détail, le serveur doit permettre à l'utilisateur d'accéder par téléphone aux informations actuellement disponibles via Internet sur les sites *citefutee.com* et *transport-idf.fr*. Ainsi, l'utilisateur aura la possibilité :

- D'indiquer oralement son trajet. Les points de départ et d'arrivée peuvent être une adresse, une gare SNCF (IDF), un arrêt de bus, de tram, de métro ou de RER, un lieu remarquable (Tour Eiffel, piscine de la Butte aux cailles, mairie de Cachan...),
- D'indiquer éventuellement le jour où il souhaite effectuer ce trajet (aujourd'hui, dimanche, le 25 novembre...)
- D'indiquer éventuellement l'heure vers laquelle il souhaite partir où arriver, (maintenant, vers 17 heures...)
- Et enfin de choisir un critère de recherche parmi les suivants : trajet le plus rapide, avec le moins de changement, avec le moins de marche à pied etc.

En réponse à la requête de l'utilisateur, le système proposera un itinéraire détaillé comprenant le réseau à emprunter (RER, train, métro, bus ou tram) la ligne sur le réseau ainsi que l'horaire théorique de départ au point d'arrêt concerné et ce pour chaque tronçon du trajet. Le temps de parcours total estimé sera également précisé.

Phases du projet : Le projet SATIM dans son intégralité est composé de huit lots avec un horizon à 30 mois **mais le financement de la PREDIM ne concerne que les lots 2 et 3** (sur 24 mois).

- Lot 2 – Collecte, analyse et transcriptions des données
- Lot 3 – Décodeur avec contraintes linguistiques

Organisation : Les actions menées sur le plan organisationnel sont les suivantes :

- Réunion de lancement le 13/01/2004 dans les locaux de Vecsys
- Mise en place d'un site web d'information (<http://www-tlp.limsi.fr/satim>) et d'un site ftp (<ftp://satim@www-tlp.limsi.fr>) permettant l'échange des documents et la collecte des données de développement.
- Mise en place de téléconférences mensuelles réunissant au moins un des représentants de chacun des quatre partenaires. Tous les comptes rendus de ces téléconférences sont disponibles sur le site internet.

Avancement :

- **Lot 2 – Collecte, analyse et transcriptions des données**

Ce lot constitue la phase I du projet SATIM. Soulignons que la collecte de données représentatives de l'application est le pré-requis essentiel pour le développement du moteur de reconnaissance, objet du lot 2. Rappelons qu'il s'agissait de collecter :

- ✓ les requêtes textuelles effectuées sur les sites *citefutee.com* (RATP) et *transport-idf.com* (AMIVIF).
- ✓ les demandes effectuées auprès des télé-conseillers du Centre Téléphonique d'Informations de la RATP.

Actions menées et état d'avancement :

- Elaboration d'un document de conventions qui définit la nature des données fournies par la RATP et leur conditions d'utilisation par le LIMSI-CNRS et VECSYS. Ce document a été validé par tous les partenaires, il est actuellement en cours de signature.
- Mise en place par la RATP d'une procédure de récupération automatique quotidienne des requêtes sur citefutee.com. Ces données sont transférées automatiquement chaque nuit sur le site ftp de SATIM.
- Normalisation et analyse par le LIMSI-CNRS des données transférées par ftp. La procédure de collecte ayant démarré le 05/07/2004, nous disposons actuellement des volumes suivants :

Nombre total de requêtes : **11 737 527**

Nombre total de requêtes exprimées par des points d'arrêt : 5 499 936

Nombre total de requêtes exprimées par des lieux remarquables : 1 083 518

Nombre total de requêtes exprimées par des adresses (noms de voiries) : 5 153 991

Nombre de requêtes distinctes sans normalisation : 3 314 337

Nombre de requêtes distinctes après normalisation : 74 126

La procédure de normalisation consiste essentiellement à unifier les requêtes concernant un même point (arrêt, lieu ou adresse) en conservant l'entrée normalisée proposée par la RATP.

Exemple :

Requête client : Vaucluse

Requête normalisée : Centre Hospitalier de Perray Vaucluse PARIS-18EME

- Le développement nécessaire à l'extraction des données de transport-idf.com a été réalisé par la SNCF et l'AMIVIF et la procédure de collecte automatique devrait être initialisée très prochainement.
- La collecte des données audio étant soumise à un accord préalable de la CNIL, la RATP a soumis un dossier décrivant le protocole d'enregistrement auprès de cet organisme. L'autorisation ayant été obtenue, Vecsys a pu installer un système d'enregistrement dans les locaux de l'Unité Information Multimédia de la RATP. La campagne d'enregistrement s'est déroulée du 15 juillet au 4 août 2004. Pendant cette période, 3000 appels ont pu être collectés. Parmi eux, seuls les appels dont la durée est supérieure à 30 secondes ont été transcrits.
- Transcription et analyse des appels : 2000 appels ont été analysés et transcrits en utilisant l'outil Transcriber. Un exemple de transcription est donné ci-dessous.

Exemple de transcription

Transcriber 1.4.2

File Edit Signal Segmentation Options Help

[conv-]Monsieur[-conv]

euh je voudrais me rendre au 45 euh 45 avenue des Etats-Unis à Versailles

(no speaker)

Agent

alors Versailles

(no speaker)

Agent

avenue des Etats-Unis

Client

ouais au 45 c'est l'IUFM euh de Versailles

(no speaker)

Agent

satim_h5_fre_1_0019

report											
Client	(no... Ag	(no speaker)	Age.	Client	f	f	Client	Agent	C	(no... Age C	
b [co r [nv]	euh je voudrais me rendre au 45... ... Versailles	al. .es		aven. ... is	ouais au 45... ... Versailles			et en fait je je suis aux. ... l'itinéraire euh	[r ... conv]	d s	(no... Age C d'ac b ...er r.

0 5 10 15 20 25 30

Cursor : 5.868 Selection : 5.868 - 8.502 (2.634)

Démarrer | Le Monde -... | satim | tap sur Ve... | WinZip (Un... | Invite de c... | 7% Transcrib... | Hummingbird Neighborhood | 16:41