



Pro novinky stiskněte dvojku?

Se standardy VoiceXML a CCXML to jde i lépe!

OptimSys, s.r.o. tel.: 541 143 065
U Vodárny 2 fax: 541 143 066
61600 Brno info@optimsys.cz
www.optimsys.cz

Představte si, že zatelefonujete na linku péče o zákazníka a automatizovaný hlasový systém vás namísto obligátního „Vítejte, pro novinky stiskněte dvojku“ uvítá slovy „Vítejte, co pro vás mohu udělat.“ Vy hlasem odpovíte „Chci znát novinky za poslední týden,“ a systém vás hned začne informovat o nejžhavějších novinkách. To není výjev ze science fiction, ale automatizovaný systém nové generace využívající technologie syntézy a rozpoznávání řeči, postavený na standardech CCXML a VoiceXML. Ve světě již poměrně běžná věc, která si pomalu nachází cestu i do České republiky.

V posledním desetiletí zasáhly oblast komunikace dva fenomény, které ji zcela změnily. Prvním z nich bylo masivní rozšíření mobilních telefonů, druhým potom stejně razantní nástup internetu a služby WWW. Webové stránky a telefonní linky péče o zákazníka se proto staly dvěma nejdůležitějšími komunikačními kanály většiny firem a institucí s vnějším světem.

Zatímco webové stránky jsou většinou z hlediska nákladů poměrně nenáročným médium s plně automatizovaným nepřetržitým provozem, náklady na vybudování a provoz call centra znamenají v rozpočtu organizací nepřehlédnutelnou položku. Přesto je nutné telefonní komunikační kanál podporovat, neboť telefon je, na rozdíl od přístupu k internetu, k dispozici prakticky vždy a prakticky všem lidem.

Logickým požadavkem provozovatele proto je co největší část telefonní komunikace automatizovat. Automatizace nejen snižuje náklady, ale přináší i výhody zákazníkům v podobě zkrácené čekací doby nebo možnosti plně automatizovaného vyřešení některých svých požadavků, a to zejména v době, kdy není linka obsluhovaná operátorem.

Nejčastější formou automatizace jsou tzv. IVR (Interactive Voice Response) systémy, které všichni známe například ze zákaznických linek bank či mobilních operátorů. V současné době převažují na trhu, zejména v České republice, IVR systémy předchozí generace, ve kterých uživatel pro získání informace musí pomocí tlačítek telefonu procházet soustavou nabídek, dokud nenalezne požadovanou informaci. Tato informace je mu poté přehrána, přičemž je složena z dopředu připravených audio nahrávek. Tyto systémy se vyznačují zpravidla malou mírou flexibility, omezenými možnostmi a vazbou na řešení konkrétního dodavatele.

CCXML + VoiceXML = hlasový server

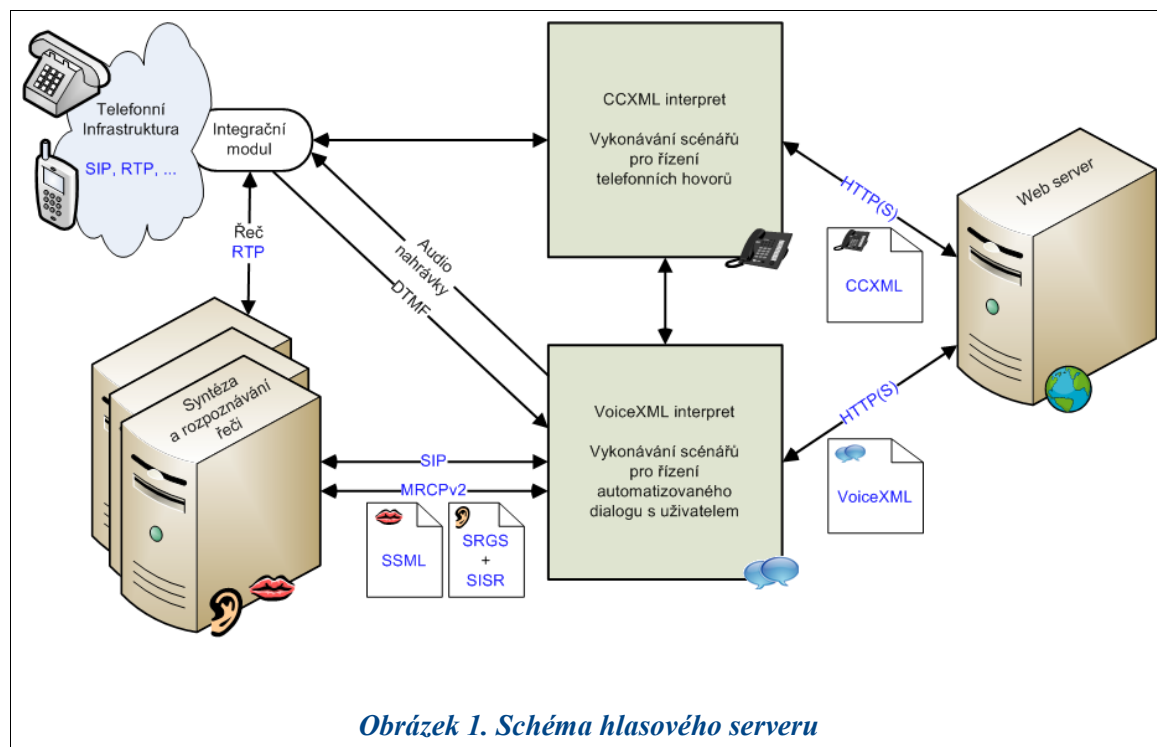
Pojetí IVR systémů se výrazně změnilo s příchodem standardů CCXML (Call Control XML) a VoiceXML. Jazyk CCXML slouží k popisu scénářů pro automatické řízení telefonních hovorů, zatímco jazyk VoiceXML je určen pro popis scénářů automatizovaného dialogu mezi člověkem a strojem.

Jejich vzájemnou integrací vznikne tzv. *hlasový server*, který je schopen volajícího automaticky obsloužit. Příkladem hlasového serveru je systém OptimTalk společnosti OptimSys. Zdlouhavé a neoblíbené procházení nabídek může být nahrazeno rozpoznáváním řeči, kdy uživatel svůj požadavek jednoduše řekne vlastními slovy a počítač mu porozumí. V případě potřeby vede počítač s uživatelem dialog za účelem doplnění a ujasnění informací. Připravené audio nahrávky mohou být nahrazeny pomocí syntézy řeči, což zvyšuje jejich flexibilitu.

Schéma hlasového serveru je znázorněno na obrázku 1.

Scénáře pro automatické řízení telefonních hovorů jsou popsány pomocí jazyka CCXML. Pomocí toho jazyka lze detailně specifikovat:

- pravidla, jak zacházet s příchozími hovory (např. přijmout/odmítnout),



- pravidla pro přesměrování hovorů na jinou destinaci (např. na operátora),
- pravidla pro spojování hovorů do konferencí,
- pravidla pro navazování odchozích hovorů
- a další.

Scénáře popsané pomocí jazyka CCXML jsou vykonávány pomocí počítačového programu zvaného *CCXML interpret*. Ten je spojen s telefonní infrastrukturou pomocí integračního modulu, který odlišuje zbytek hlasového serveru od technických specifik a komunikačních protokolů této telefonní infrastruktury. V prostředích využívajících VoIP telefonii komunikuje integrační modul s telefonní infrastrukturou zpravidla pomocí standardizovaného protokolu SIP, hlasové proudy jsou pak přenášeny pomocí standardizovaného protokolu RTP.

Jedna z operací definovaná v jazyce CCXML je napojení telefonního hovoru na automatizovaný systém komunikující s uživatelem. Příslušný komunikační scénář je popsán pomocí jazyka VoiceXML. Pomocí tohoto jazyka lze popsat informace, které je potřeba od uživatele získat, a způsob, jak tyto informace získat. VoiceXML scénáře jsou vykonávány pomocí počítačového programu zvaného *VoiceXML interpret*.

Pro vytváření VoiceXML scénářů lze využít mimo jiné tyto základní operace:

- zpracování mluveného vstupu uživatele pomocí rozpoznávání řeči,
- zpracování vstupu uživatele zadávaného prostřednictvím tlačítek telefonu pomocí tónové volby (DTMF),
- nahrávání mluveného vstupu uživatele,
- přehrávání audio nahrávek uživateli,

- přehrávání zpráv uživateli vytvořených z textu pomocí syntézy řeči
- a další.

Největší síla jazyka VoiceXML spočívá v možnosti vést s uživatelem dialog a přizpůsobovat svou komunikační strategii aktuální situaci. Například nesdělí-li uživatel počítači všechny požadované informace, počítač se na ně v dalších krocích postupně dotáže. Naopak má uživatel možnost v jednom kroku sdělit počítači více informací, než je po něm požadováno, a zkrátit tak konverzaci. Rovněž pořadí, ve kterém jsou od uživatele informace získávány, není pevně dáno, a může se měnit.

Další standardy ve hře

Dalším důležitým standardem pro hlasový server je Media Resource Control Protocol v2 (MRCPv2). Ten slouží pro klient-server komunikaci mezi VoiceXML interpretem a řečovými syntetizéry a rozpoznávači umístěnými na serverech v síti. Tato distribuovaná architektura je nezbytná z důvodu škálovatelnosti, neboť řečové rozpoznávače a syntetizéry jsou obecně náročné na výpočetní a paměťové zdroje.

Pro přenos hlasu do rozpoznávače a ze syntetizéru slouží opět protokol RTP, pro inicializaci relace mezi VoiceXML interpretem a MRCPv2 serverem se využívá již zmíněný protokol SIP.

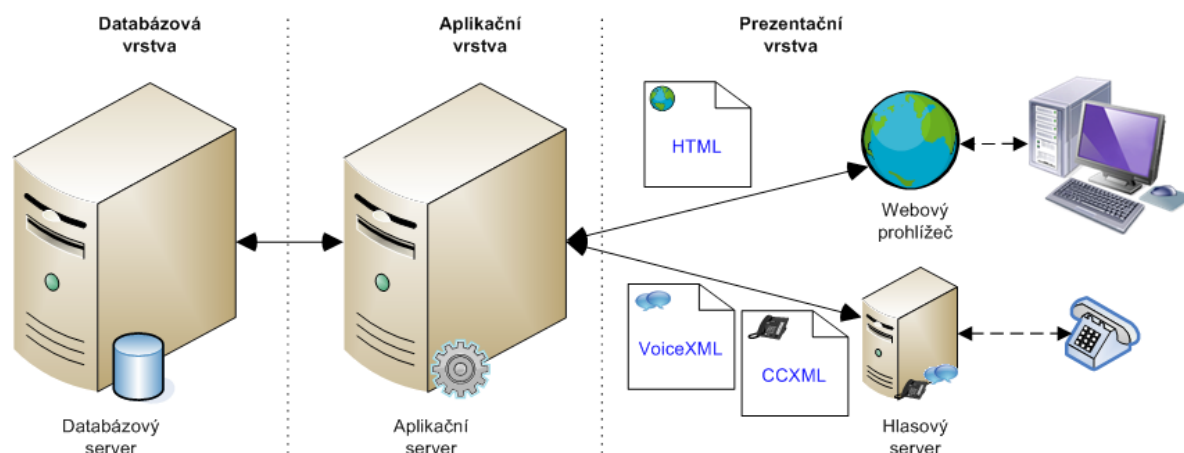
Řečovému syntetizéru se mimo jiné posílá text, který má být převeden na mluvenou řeč. Tento text je zpravidla doplněn značkami ovlivňujícími výstup syntetizéru (např. použitý hlas, intonace, tempo řeči, pauzy v řeči apod.). Pro toto značkování se používá standard Speech Synthesis Markup Language (SSML).

Rozpoznávači řeči se mimo jiné posílá definice očekávaných frází, které může uživatel říci, ve formě tzv. *gramatiky*. Formát této gramatiky je definován standardem Speech Recognition Grammar Specification (SRGS). Gramatika může být dále obohacena o informace popisující, jak k dané vyřčené frázi přiřadit reprezentaci jejího významu, neboť teprve význam je informace, která je potřebná pro vedení dialogu. Například vyřčené fráze „Ano“, „Jistě“ a „Přesně tak“ mají typicky stejný význam, neboť všechny reprezentují souhlas. Tento proces vyhodnocování významu fráze a formát anotace v rámci gramatik je definován standardem Semantic Interpretation for Speech Recognition (SISR).

Normy CCXML, VoiceXML, SSML, SRGS a SISR jsou standardizovány konsorciem W3C a dohromady tvoří rámec, jenž se nazývá *W3C Speech Interface Framework*. Normy SIP, RTP a MRCPv2 jsou standardizovány organizací IETF. Ve všech případech se jedná o otevřené a patenty nezátížené standardy, jejichž specifikace jsou volně dostupné. V případě CCXML a MRCPv2 nebyl ještě proces standardizace ukončen, lze se však již běžně setkat s jejich komerčními implementacemi, včetně implementace společnosti OptimSys.

Integrace hlasového serveru s existujícími systémy

Je zřejmé, že oba výše zmíněné komunikační kanály, tedy webová prezentace i telefonní linka péče o zákazníka, musí pracovat se stejnými daty a se stejnou vnitřní logikou, pouze forma prezentace informací uživateli je odlišná. Technologie VoiceXML a CCXML byly navrženy tak, aby bylo možné hlasový server snadno integrovat do standardní třívrstvé architektury, která tvoří základ většiny moderních informačních systémů a webových aplikací. Schéma této integrace je zachyceno na obrázku 2.



Obrázek 2. Hlasový server v třívrstvé architektuře

Běžně uživatel zasílá požadavky a data aplikačnímu serveru pomocí webového prohlížeče, kterému aplikační server zasílá odpovědi zpět ve formě HTML stránek. Tyto stránky pak webový prohlížeč graficky zobrazuje uživateli. Při použití hlasového serveru je jediným rozdílem to, že uživatel místo klávesnice a monitoru používá telefon, roli webového prohlížeče supluje hlasový server, odpovědi z aplikačního serveru jsou zasílány hlasovému serveru pomocí jazyků CCXML a VoiceXML a uživateli jsou prezentovány hlasově.

Integrace hlasového serveru do webové infrastruktury zákazníka nevyžaduje žádné zásadní změny v existujících systémech a aplikacích. Hlasový server „pouze“ možnosti systému rozšiřuje o nový komunikační kanál a případně nové aplikace.

Komunikace hlasového serveru s aplikačním serverem probíhá pomocí standardizovaného protokolu HTTP. V případě potřeby šifrovaného spojení lze použít protokol HTTPS. Z VoiceXML i CCXML scénářů lze snadno volat webové služby pomocí standardizovaného protokolu SOAP. Do CCXML interpretu je možné posílat požadavky z externích aplikací rovněž pomocí HTTP protokolu a stejně tak z něj požadavky přijímat.

Hlasový server může tedy rovněž sloužit jako výkonná a všestranná platforma pro integraci počítačových systémů s telefonii (CTI, computer telephony integration).

Výhody CCXML a VoiceXML

Použití CCXML, VoiceXML a souvisejících technologií popsaných výše přináší ve srovnání s technologiemi dřívějších generací dramatické snížení nákladů na implementaci řešení (až o 90 %) a zkrácení času pro jejich uvedení na trh (o více než 2/3) [1]. Výsledná řešení jsou vysoce flexibilní, robustní, snadno udržovatelná a modifikovatelná. Díky orientaci na prověřené webové standardy jsou rovněž v maximální možné míře využívány existující zdroje, neboť webová infrastruktura je dnes přítomná prakticky v každé organizaci.

Automatizované IVR transakce rovněž výrazně snižují provozní náklady. Podle [2] se cena jedné automaticky obslužené transakce pohybuje v rozmezí 0,05 – 0,40 USD, zatímco cena transakce obslužené operátorem v rozmezí 3,50 – 6,00 USD (údaje pro americký trh).

Ve prospěch těchto technologií hovoří také fakt, že se jedná o otevřené, mezinárodní, patenty nezařízené standardy. Tím je zajištěna zejména interoperabilita mezi jednotlivými částmi řešení od různých dodavatelů a to i vzhledem k technologiím vyvinutým v budoucnu. Otevřené standardy dále přináší širší možnosti automatizace a zaručují existenci širokého spektra nástrojů od mnoha výrobců. Rovněž jsou zárukou větší dostupnosti školených profesionálů schopných tvořit a udržovat řešení založená na těchto standardech. V neposlední řadě je potřeba zmínit dobrou dostupnost kvalitní dokumentace a literatury.

Použití technologií CCXML a VoiceXML tedy snižuje celkové náklady na vlastnictví a znamená výrazně nižší riziko a lepší ochranu investice.

Tabulka 1 shrnuje hlavní rozdíly mezi předchozí a novou generací automatizovaných telefonních a hlasových systémů.

<i>Předchozí generace</i>	<i>Nová generace</i>
Založená na proprietárních technologiích	Založená na otevřených mezinárodních standardech CCXML a VoiceXML
Ovládání pomocí tlačítek telefonu	Ovládání pomocí tlačítek telefonu a rozpoznávání mluvené řeči
Přehrávání předem připravených nahrávek	Přehrávání předem připravených nahrávek a syntéza mluvené řeči
Základní podpora Voice over IP (VoIP)	Snadná integrace s Voice over IP (VoIP)
Tvorba aplikací pomocí proprietárních nástrojů	Tvorba aplikací podobná návrhu HTML stránek, dostupné nástroje různých výrobců

Tabulka 1: Srovnání předchozí a nové generace automatizovaných telefonních a hlasových systémů

Aplikační oblasti

Použití technologií CCXML a VoiceXML otevírá zcela nové oblasti pro aplikace přístupné přes telefon. Například automatizace vyhledávání dopravních spojení přes telefon byla pomocí technologií předchozí generace těžko realizovatelná, neboť zadání jména lokace pomocí tlačítek telefonu je velmi těžkopádné. Pomocí rozpoznávání řeči lze tento problém elegantně vyřešit. Jiným příkladem může být systém pro čtení mailů, který lze realizovat pomocí syntézy řeči.

Jednou z nejběžnějších aplikačních oblastí je tvorba automatizovaných informačních a rezervačních linek. Typickými zástupci jsou linky poskytující informace o dopravní situaci, jízdních řádech, počasí, kultuře, sportu či turistické informace, dále linky pro rezervaci letenek, hotelu, auta, systémy pro objednávky zboží a sledování zásilek, linky pro bankovní a pojišťovací služby, linky telekomunikačních operátorů nebo linky s informacemi o produktech.

Velké uplatnění nalézají tyto technologie v call centrech, neboť dokáží automaticky obsloužit jednoduché, často se opakující požadavky, a operátoři se tak mohou věnovat složitějším problémům. Různé zdroje uvádějí míru automatizace od 40 do 90 %. Hlasový server může tvořit i samotné srdce call centra řídicí obsluhu telefonních hovorů včetně použití algoritmů automatické distribuce hovorů, prediktivního vytáčení apod.

Velkým přínosem je automatizace komunikace při řešení krizových situací, jako jsou požáry, úniky nebezpečných látek nebo jiné bezpečnostní incidenty v objektech či areálech. Technologie syntézy řeči spolu s možností automatického řízení komunikačních linek umožňuje automatizovat komunikační body krizového plánu. Na rozdíl od člověka může počítačový systém komunikovat se všemi zainteresovanými lidmi naráz, čímž výrazně uspoří čas. To může být rozhodující pro záchranu lidských životů či omezení ekonomických ztrát. Zároveň eliminuje riziko provedení chyby ze strany obsluhy, jež se zvyšuje vlivem stresu. V praxi může být výstupem pro obsluhu, která v kritickém čase obstarává ostatní činnosti, souhrnná zpráva o proběhlé komunikaci. Spolehlivost doručení automatizovaných zpráv lze zajistit formou zadání potvrzovacího kódu příjemcem pomocí tlačítek telefonu.

Související oblastí je vzdálený monitoring a správa zařízení. Nekritické monitorovací systémy často posílají SMS na zadané číslo v případě, že je detekováno vybočení sledovaných veličin z nastavených limitů. Nahrazení SMS telefonátem s možností ověřit příjem zprávy zvyšuje spolehlivost a efektivitu těchto systémů. Je rovněž možné, aby si uživatel zavolal na definované telefonní číslo, kde jsou mu pomocí syntézy řeči sděleny potřebné informace. Navíc lze uživateli nabídnout možnost zapnutí, vypnutí nebo změny konfigurace zařízení po telefonu. Tyto systémy najdou uplatnění v moderních domech, datových centrech a jiných monitorovaných provozech.

Technologie VoiceXML a CCXML jsou vhodné rovněž pro tvorbu systémů uživatelské kolaborace a unifikovaného zpracování zpráv zahrnující aplikace hlasové pošty, čtení emailů pomocí řečového syntetizéru, plánování schůzek a upozorňování na blížící se schůzky, automatické vytváření konferenčních hovorů atd.

Zajímavou aplikací je automatická spojovatelka/recepční. Aplikace může sloužit k automatickému spojování příchozích i interních hovorů v rámci organizace. Systémy rozpoznávání řeči jsou již na takové úrovni, že dokáží rozpoznávat seznamy jmen čítající i desítky tisíc položek (pro angličtinu i řádově více). Systém je navíc schopen vést s uživatelem dialog za účelem upřesnění nejednoznačností. Aplikace může zajišťovat i další úkony jako je podávání informací nebo přebírání vzkazů. Při napojení na interní docházkový systém obsahující informace o přítomnosti, nepřítomnosti a aktivitách jednotlivých zaměstnanců může systém automaticky rozhodnout, zda má příchozí hovor směřovat na telefon do kanceláře, mobilní telefon nebo má raději převzít vzkaz.

Hlasový server může rovněž sloužit jako výkonná a všestranná platforma pro integraci počítačových systémů s telefoníí (CTI, computer telephony integration). Typickým příkladem je integrace se systémy řízení vztahu se zákazníky (CRM). Snadno lze do CRM systému implementovat funkci vytáčení na kliknutí využívající k navázání hovoru hlasový server. Ten může rovněž zaslat CRM systému notifikaci v případě příchozího hovoru, takže CRM systém může uživateli zobrazit informace o volajícím ještě dříve, než uživatel hovor přijme.

Potenciál technologií VoiceXML a CCXML byl naplno využit společností OptimSys při implementaci systému OptimCall. Tento systém pracuje jako nadstavba pobočkové telefonní ústředny a umožňuje uživateli nastavit chování své telefonní linky bez nutnosti zasahovat do konfigurace ústředny. Zatímco pobočková ústředna často nabízí pouze několik základních funkcí pro zpracování hovorů, OptimCall nabízí podstatně vyšší flexibilitu a detailní možnosti nastavení pro každou linku. Příchozí hovory je možné zpracovat následovně: nechat hovor zvonit na volanou linku (žádná akce); nechat hovor zvonit na několik čísel současně, popřípadě začít na některých z nich zvonit se zpožděním; přeměřovat hovor na jiné číslo; přeměřovat hovor do hlasové schránky, nahraný vzkaz může být poslán emailem nebo zpřístupněn přes webové rozhraní; přeměřovat hovor na automatický hlasový systém (IVR); hovor ignorovat nebo odmítnout; po zvednutí uživatelem hovor nahrávat a další. Různé akce mohou být na příchozí hovor aplikovány na základě aktuálního času

a telefonního čísla volajícího, přičemž různé sady pravidel mohou být aplikovány na pracovní dny, víkendy nebo pro případ, že je uživatel na schůzce, na služební cestě nebo na dovolené.

Aplikací však lze nalézt daleko více. Z dalších aplikačních oblastí jmenujme alespoň systémy pro telefonický průzkum, možnost rozšíření systémů pro podporu zákazníků (help desk systémů) nebo systémy pro automatickou distribuci hlášení a oznámení.

Jsou opravdu tak dobré?

Technologie CCXML a VoiceXML zcela mění pojetí automatizovaných telefonních a hlasových systémů. Reflektují požadavky současného trhu a přináší do této oblasti potřebnou flexibilitu, interoperabilitu a standardizaci. Díky využití moderních technologií, včetně technologie syntézy a rozpoznávání řeči, umožňují vyšší stupeň automatizace komunikace než tradiční IVR systémy a zpřístupňují pro automatizaci zcela nové aplikační oblasti.

Podle studie společnosti Datamonitor [3] převýšil objem dodávek IVR systémů založených na VoiceXML v roce 2008 poprvé objem dodávek tradičních IVR systémů. Podle studie společnosti T3i Group [4] bude v roce 2013 95 % všech dodávaných IVR systémů podporovat VoiceXML a 90 % systémů bude připojeno do telefonní sítě prostřednictvím VoIP telefonie na protokolu SIP. Tyto projekce jasně ukazují, že použití standardů CCXML a VoiceXML pro tvorbu moderních automatizovaných telefonních a hlasových aplikací je správnou volbou.

Reference

- [1] E. Jackson: Speaking Up For Cost Savings In The Call Center: VoiceXML Takes On The Dinosaur Of Legacy IVR. Customer Interaction Solutions Magazine, August 2003.
- [2] Donna Fluss: Why Hosted IVR May Be Right For You. CRMxchange, July 2009.
- [3] Datamonitor: Leading Speech Applications That Will Unlock Enterprise Budgets (Strategic Focus). October 2008.
- [4] T3i Group: InfoTrack for Converged Applications 2008 IVR Market Report. July 2009.