

VoiceXML

Rakenteiset dokumentit (TJTD60) – Harjoitustyö

Avainsanat: Rakenteiset dokumentit, TJTD60, VoiceXML

Janne Heinonen

Jyväskylän yliopisto
Tietojenkäsittelytieteiden laitos
Suomi
jatahein@cc.jyu.fi

Tuomas Yli-Olli

Jyväskylän yliopisto
Tietojenkäsittelytieteiden laitos
Suomi
tuomas@cc.jyu.fi

Tiivistelmä

Tämä Rakenteisten dokumenttien (TJTD60)–kurssin harjoitustyö esittelee äänikäyttöliittymien laatimiseen kehitettyä VoiceXML-kieltä ja kertoo miten ja mihin sitä käytetään.

Sisältö

1. Johdanto

- 1.1 Toimeksianto ja työn tavoite
- 1.2 Työn kappaleiden sisältö
- 1.3 Äänikäyttöliittymistä
- 1.4 Tausta, käsitteet ja määrittelyt

2. VoiceXML ja liitännäiskielet

- 2.1 VoiceXML:n esittely
 - 2.1.1 VoiceXML sovelluksen rakenne
 - 2.1.2 VoiceXML-sovelluksen toiminnallisuus
 - 2.1.3 VoiceXML-dokumentin rakenne

- 2.1.4 VoiceXML-dokumentin toiminnallisuus
 - 2.1.4.1 Alidialogit
 - 2.1.4.2 Tapahtumat
- 2.2 Palvelinpuolen ratkaisut ja toteutusarkkitehtuuri
 - 2.2.1 Arkkitehtuurimalli
- 2.3 Oheiskielet
 - 2.3.1 SRGS – The Speech Recognition Grammar Specification
 - 2.3.2 N-Gram specification
 - 2.3.3 SSML – The speech synthesis specification
 - 2.3.4 SISR – Semantic Interpretation for Speech Recognition
 - 2.3.5 PLS – Pronunciation Lexicon Specification
 - 2.3.6 CCXML – Call Control eXtensible Markup Language
 - 2.3.7 SCXML – State Chart XML: State Machine Notation for Control Abstraction
- 2.4 Kilpailijat
 - 2.4.1 SALT

3. VoiceXML käytännössä

- 3.1 Esimerkkejä VoiceXML:n käyttöönotosta
 - 3.1.1 E*TRADE FINANCIAL
 - 3.1.2 Spain's Bankinter
- 3.2 Demonstraatio

4. Yhteenveto

- 4.1 Loppuyhteenveto
- 4.2 Tenttikysymys

Lähteet

1. Johdanto

1.1 Toimeksianto ja työn tavoite

Tässä harjoitustyössä esittelemme äänikäyttöliittymien laatimiseen kehitetyn VoiceXML-kielen **[VoiceXML, 2004]** ja kerromme miten ja mihin sitä käytetään. Listaamme sovellusalueita ja kerromme muutamista todellisista käyttötapauksista. Lisäksi käymme läpi VoiceXML:ään oleellisena osana kuuluvat liitännäiskielet, kuten SSML (Speech Synthesis Markup Language) **[SSML, 2004]**. Kartoitamme myös kilpailevat teknologiat lyhyesti. Harjoitustyöhön liittyy myös malli puheen merkkauksesta: "Runonlausunta–automaatti".

1.2 Työn kappaleiden sisältö

Ensimmäisessä kappaleessa esitellään äänikäyttöliittymiä ja niiden käyttöä yleisellä tasolla. Kappaleessa 2 perehdytään VoiceXML tekniikkaan, sen toteutusarkkitehtuuriin, sekä olennaisena osan mukaan kuuluviin liitännäiskieliin. Kappaleessa 3 käydään läpi VoiceXML:n sovellusmahdollisuuksia sekä esitellään kaksi käytännön toteutusesimerkkiä. Kappale 4 kokoaa raportin oleellisen sisällön yhteen sekä tarjoaa suppean katsauksen tulevaan.

1.3 Äänikäyttöliittymistä

Äänikäyttöliittymät tarjoavat vaihtoehtoisen tavan ihmisen ja tietokoneen väliseen kommunikaatioon. Äänen käyttö voi olla myös osa graafista käyttöliittymää, onhan visuaalinen ja auditorinen systeemi ihmisillä kehittynyt toimimaan yhdessä, toisiaan tukien. Esimerkiksi sulautetuissa järjestelmissä ja mobiililaitteissa on hyvin pienet ja rajalliset näytöt. Tähän ongelmaan on eräs ratkaisu se, että säästetään rajallista näyttötilaa esittämällä dataa äänellä. Äänikäyttöliittymiä puolustetaan myös näkökulmasta, jonka mukaan nykyiset näyttöihin perustuvat käyttöliittymät käyttävät ihmisen visuaalista havaintokykyä liiankin intensiivisesti. Käyttäjän kuormitusta voidaan vähentää esittämällä tärkeää informaatiota äänen avulla, jolloin havainto jakaantuu useammalle aistille.

Ääni on voimakas elementti käyttöliittymässä—Siinä missä näkö tarjoaa vain kapean kaistaleen ympäröivästä maailmasta, yltää kuulo puolestaan joka puolelle. Sekä näkö— että kuuloaisti aiheuttavat orientaatioefektin ympäristön tapahtumiin, mutta katse on helppo kääntää pois kohteesta, tahallaan tai epähuomiossa. Ääntä on vaikeampi välttää. **[Benyon, 2005]**

Puhepohjaiset järjestelmät palvelevat erityisesti:

- Kun käyttäjällä on näköongelmia tai katse on suunnattava poispäin laitteesta käyttötilanteesta johtuen
- Kun käyttäjän kädet ovat varatut tai käyttöolosuhteet estävät näppäimistöllä tapahtuvan syötön
- Liikkuvassa / mobiilissa käyttötilanteessa

Puhetta ei ole aiemmin käytetty laajasti käyttöliittymissä lukuun ottamatta erikoistarpeita. Yksi syy on se, että äänikäyttöliittymä sopii vain tiettyihin ympäristöihin. Graafisella käyttöliittymällä on myös selkeät vahvuutensa ääntä vastaan: Virheilmoitukset pysyvät selkeästi esillä, tilatieto on esillä, samoin kuin valintavaihtoehdot menujen ja nappuloiden muodossa. Vastaavasti audiopohjaiset käyttöliittymät ovat hetkellisiä. Näin ollen ne eivät voi olla ainakaan kerta—annoksina kovin laajoja valintamahdollisuuksiensa suhteen. Vaikka äänipohjaisia palveluita voidaan tehdä varsin kattavasti nykytekniikalla, ei kaikkea ole silti välttämättä järkevää toteuttaa. Äänipohjaisilla systeemeillä onkin omat käyttötilanteensa ja erikoispiirteensä, jotka Shneiderman ja Plaisant (2005) ovat hausalla analogialla kiteyttäneet: "Speech is the bicycle of user—interface design: It is great fun to use and has an important role, but it can carry only a light load. Sober advocates know that it will be tough to replace the automobile, graphical user interfaces." **[Shneiderman, 2005]**

Esteitä puheentunnistusteknologian käytölle:

- Meluisa käyttöympäristö
- Epävarma tunnistaminen, kun käyttäjät ja käyttöympäristö vaihtelevat

Puhemuotoisen sisällön haittoja:

- Hidas saanti verrattuna lukemiselle ja katsomiselle
- Puheen hetkellinen luonne
- Hakutoimintojen hankaluus verrattuna tekstiin

Tietyistä haittapuolista ja rajoitteista huolimatta äänikäyttöliittymien saralla nähdään paljon mahdollisuuksia ohjelmisto- ja palvelukehittäjien keskuudessa. Varsinkin standardit äänen rakenteiselle merkkaukselle verkkoympäristöstä (kuten VoiceXML ja SALT) ja jatkuvasti kehittyvät puheen tuottamis- ja tunnistamisteknologiat mahdollistavat uusia innovatiivisia sovelluksia. Uusia mahdollisuuksia tarjoavat visuaalisia ja auditivisia elementtejä yhdistelevien käyttöliittymien lisäksi esimerkiksi puhelimen yhteyteen rakennetut palvelut, jolloin puhelimella on mahdollista päästä käsiksi verkkosisältöihin äänivälitteisesti. Tämä mahdollistaisi tietyssä mittakaavassa myös sen, ainakin periaatteessa, että kuka tahansa pääsisi käsiksi verkon sisältöihin mistä tahansa puhelimesta.

1.4 Tausta, käsitteet ja määritykset

VoiceXML [**VoiceXML, 2004**] on nimi teknologiastandardille, jonka kehittämisestä ja hallinnoinnista vastaa VoiceXML Forum. VoiceXML on pitkälle kehitetty formaalikieli, jonka avulla voidaan laatia loogisia puurakenteita ja tehdä verkkopalveluita esimerkiksi puhelimella käytettäväksi. XML-pohjainen VoiceXML perustuu aiemmin kehitettyihin tekniikoihin kuten VoXML:ään (Motorola [**VoxML, 1999**]) ja SpeechML:ään (IBM [**Coverpages, 1999**]). Kyseessä on yksi standardiehdokas puheääntä käyttäviin käyttöliittymiin. VoiceXML on nuori teknologia – määrittelyn versio 1.0 jätettiin maaliskuussa 2000 hyväksyttäväksi. Uusin versio on tällä hetkellä 2.1 ja se ajoittuu kesäkuulle 2005.

VoiceXML Forumin ovat perustaneet AT&T, Lucent Technologies, Motorola, ja IBM tarkoituksenaan "aim to drive the market for voice- and phone-enabled Internet access by promoting a standard specification for VoiceXML, a computer language used to create Web content and services that can be accessed by phone."

Langattomat viestintävälineet kärsivät pienistä näytöistä, rajallisesta tiedonsyöttökäytöstä ja pienestä prosessoritehosta. Kukaan ei kuitenkaan kyseenalaista niitä perinteisen puheviestinnän välineinä, mutta nähtäväksi jää, ottaako suuri yleisö ne myös muunlaisen tiedon siirtovälineiksi. Yksi vaihtoehto tekstipohjaisille käyttöliittymille (kuten esim WAP) on alun perin IVR:nä (Interactive Voice Response) tunnettu järjestelmä. [**WDN, 2005**]

IVR on tietokonepohjainen järjestelmä, jonka avulla puhelimella soittava käyttäjä voi valita vaihtoehtoja äänivalikoista ja ohjailla tietokonetta. Yleensä järjestelmä toistaa ennalta äänitettyjä puhepätkiä, joihin käyttäjä vastaa puhelimen näppäimillä tai yksinkertaisella puheella, kuten "Kyllä /

Ei" tai jokin numero. Tällaisia järjestelmä käytetään esimerkiksi puhelinpankeissa, lentojen varauspalveluissa tai soittajan ohjailussa edemmäs palveluihin, kuten monissa call centerissä tehdään. IVR –ohjatun puhepalvelun vuo voidaan laatia vaihtelevin tavoin. Vanhempia järjestelmiä on ohjelmoitu erityisillä ohjelmointi- ja skriptauskielillä, uudemmat järjestelmät toteutetaan useimmiten samantyyllisesti kuin WWW–sivut. Tällöin käytetään esimerkiksi VoiceXML tai SALT-kieliä. Kehitystyö on näiden välineiden avulla useampien ulottuvissa, ja periaatteessa kenellä tahansa web-kehittäjällä on tarvittavat työkalut puheluvuon laatimiseksi. [DTSI, 2002]

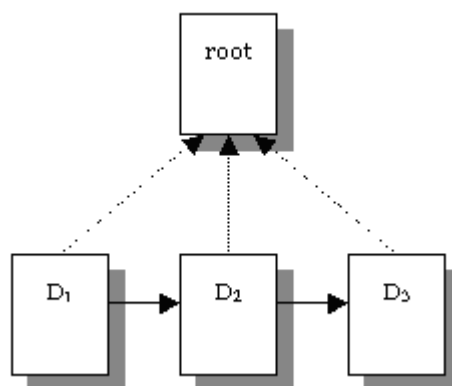
2. VoiceXML ja liitännäiskielet

2.1 VoiceXML:n esittely

VoiceXML on XML-pohjainen merkintäkieli, joka on suunniteltu toteuttamaan audiodialogeja, joissa voidaan käyttää synteettistä puhetta, digitalisoitua audiota, tunnistaa puhe- sekä DTMF (Dual Tone Multifrequency) näppäinkomentoja, tallentaa puhuttu syöte sekä käydä keskustelua, jossa käyttäjän puhetta kuunnellaan ja siihen reagoidaan sovelluksen tasolta. Näiden lisäksi VoiceXML:ään liittyy perinteisen puhelikeskuksen toimintoihin kuuluvia ominaisuuksia kuten puheluun vastaaminen, sen katkaisu sekä siirto.

2.1.1 VoiceXML sovelluksen rakenne

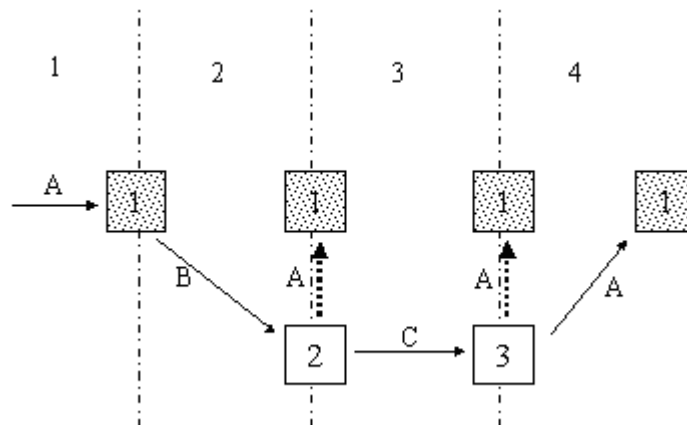
Käyttäjän yhtäjaksoista vuorovaikutusta VoiceXML toteutuksen kanssa sanotaan istunnoksi (session). Istunnon aikana voidaan liikkua, joko yhden tai useamman VoiceXML sovelluksen sisällä. VoiceXML sovellus puolestaan koostuu joukosta VoiceXML-dokumentteja, joilla on sama juuridokumentti. Muita sovelluksen dokumentteja kutsutaan lehtidokumentiksi. Sovelluksen rakenne havainnollistuu kuvasta 1. Kuvassa on merkitty nelisymboleilla VoiceXML dokumentteja. Nuolilla kuvataan smiten dokumentista voi siirtyä toiseen.



Kuva 1: VoiceXML sovelluksen dokumenttirakenne

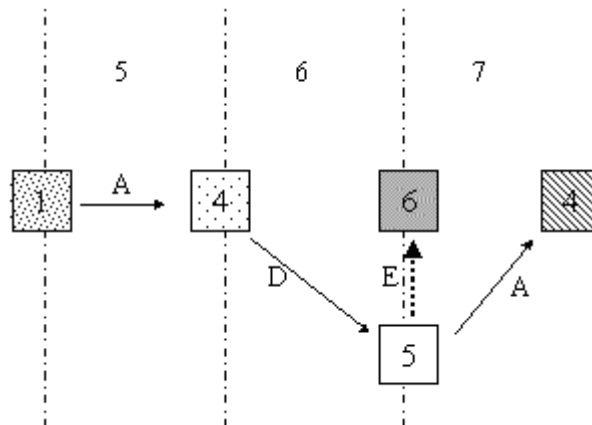
2.1.2 VoiceXML-sovelluksen toiminnallisuus

Sovelluksen juuridokumentti ladataan aina samalla muistiin kun joku sovelluksen lehtidokumenteista ladataan. Mikäli sovelluksen juuridokumentista siirrytään sen lehtidokumenttiin, pidetään dokumentti muistissa. Näin sovelluksen yhteisiä muuttujia ja kielioppia (myös muita ominaisuuksia) voidaan pitää juuridokumentissa jolloin ne ovat myös kaikkien lehtidokumenttien käytettävissä. Kuva 2 havainnollistaa mitkä dokumentit ovat ladattuna tietynä sovelluksen suoritushetkellä. Kuvassa on merkitty neliöillä VoiceXML sovelluksen dokumentit. Tummat neliöt esittävät sovelluksen juuridokumenttia ja vaaleat lehtidokumentteja. Pystyviivat kertovat mitkä sovelluksen dokumentit ovat ladattuna tietynä hetkenä.



Kuva 2: VoiceXML sovelluksen sisällä liikkuminen

Sovelluksen suoritus loppuu kun sovelluksen sen hetkinen dokumentti ei määrää enää uutta dokumenttia johon mennä tai kun siirrytään dokumenttiin, joka on joko toisen sovelluksen juuridokumenttia tai lehtidokumentti. Kuva 3 havainnollistaa miten yhden VoiceXML istunnon aikana voi siirtyä sovelluksesta toiseen ja mitkä sovellusten dokumentit ovat ladattuna tietynä sovelluksen suoritushetkellä. Kuvassa on merkitty neliöillä VoiceXML sovelluksen dokumentit. Tummat neliöt esittävät sovellusten juuridokumenttia ja vaalea lehtidokumenttia. Lehtidokumentti kuuluu kuvassa siihen sovellukseen, jonka juuridokumentti on sen yläpuolella. Numero neliön sisällä yksilöi dokumentit. Pystyviivat kertovat mitkä sovellusten dokumentit ovat ladattuina tietynä hetkenä.



Kuva 3: VoiceXML sovelluksesta toiseen siirtyminen

2.1.3 VoiceXML-dokumentin rakenne

VoiceXML-rakenne on määritelty DTD-määrittelyllä. [VoiceXML DTD, 2004] Seuraavassa VoiceXML-koodiesimerkissä näytetään miten perinteinen "Hello World!" merkittäisiin VoiceXML-rakenteen mukaisesti. [DTSI, 2002]

```
<?xml version='1.0'?>
<vxml version="1.0">
  <form>
    <block>Hello World!</block>
  </form>
</vxml>
```

VoiceXML:n juurielementti on <vxml>. Tärkeimmät elementit ovat juurielementti dialogielementit <form> ja <menu>. Muita toiminnallisuuden kannalta tärkeitä elementtejä ovat dokumentissa siirtymisen mahdollistavat <choice>, <goto>, <link> ja <submit>-elementit, alidialogin mahdollistava <subdialog> elementti, tapahtumiin liittyvät <catch> ja <throw> elementit sekä muuttujia määrittelevä <var> elementti.

VoiceXML-dokumentin toiminnallisuus liittyy pääasiassa dialogielementteihin, joihin muut edellä luetellut elementit vain lisäävät toiminnallisuutta. Yksi dokumentti voi sisältää useita dialogeja. Dialogeja käytetään vuorovaikutuksen luomiseen käyttäjän kanssa. Lomaketta <form> käytetään sekä informaation esittämiseen <block> ja <prompt> että informaation keräämiseen <field> käyttäjältä. Valikko <menu> voidaan mieltää erityiseksi lomakkeeksi, joka antaa käyttäjälle ennalta määrätty vaihtoehdot <choice>, jonka määräämiä polkua pitkin dialogi etenee seuraavaan dialogiin. VoiceXML dokumentti voidaan ajatella dialogeista koostuvana tilakoneena. Käyttäjä on kuitenkin yhtenä hetkenä yhdessä ja vain yhdessä dialogissa kerrallaan. Dialogi huomioi käyttäjän syötteet joiden mukaan sitten siirrytään seuraavaan dialogiin. Dialogista toiseen siirtyminen tapahtuu URI-osoitteiden avulla. Seuraava VoiceXML-koodiesimerkki kuvaa miten runokoneen toiminnallisuus

den voisi rakentaa VoiceXML-sovelluksella. Runokone on automaatti, jossa voi valita haluamansa runon, jonka jälkeen runokone lausuu kyseisen runon. Kun runo on lausuttu palataan takaisin alkuvalikkoon

Sovelluksen juuridokumentti (runoautomaatti.vxml)

```
<?xml version="1.0" encoding="UTF-8"?>
<vxml version="2.0" xmlns="http://www.w3.org/2001/vxml"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://www.w3.org/2001/vxml
  http://www.w3.org/TR/voicexml20/vxml.xsd">
<menu>
  <prompt>
    Tervetuloa runokokoelmaan. Valitse runosi: <enumerate/>
  </prompt>
  <choice next="pinkku.vxml">
    Pingiviini pakastimessa
  </choice>
  <choice next="runo2.vxml">
    Toinen runo
  </choice>
  <choice next="runo3.vxml">
    Kolmas runo
  </choice>
  <choice>
    Lopeta
    <exit/>
  </choice>
  <noinput>Ole hyvä ja valitse jokin seuraavista
vaihtoehtoista<enumerate/></noinput>
</menu>
</vxml>
```

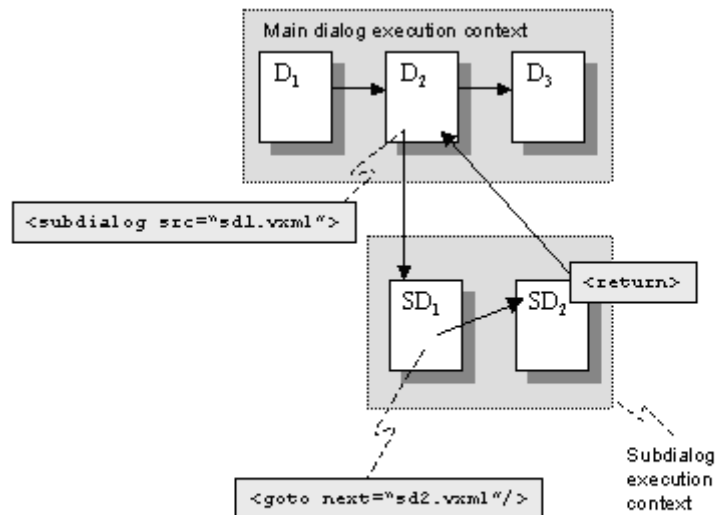
Esimerkki yhdestä alidokumentista: Leaf document (pinkku.vxml)

```
<?xml version="1.0" encoding="UTF-8"?>
<vxml xmlns="http://www.w3.org/2001/vxml"
      xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
      xsi:schemaLocation="http://www.w3.org/2001/vxml
      http://www.w3.org/TR/voicexml20/vxml.xsd"
      version="2.0" application="runoautomaatti.vxml">
  <form>
    <prompt>
<voice gender = "male" category = "adult">
  Meillä on pingviini pakastimessa. Mikäs sen on siellä lekotellessa, asia kun
on sillä tavalla, että siellä on kylmä kuin navalla. Jos tahdot pingviinin
tavat, sinun täytyy pakastin avata. Lintu sinut säikyttää ja sekamehut
silmille läikyttää. Se tuntuu kipeältä kun se mehu on aivan jäässä ja siitä
tulee otsaan patti, suuri kuin lehmän tatti.
</voice>
    </prompt>
    <block>
      <goto next="runoautomaatti.vxml"/>
    </block>
  </form>
</vxml>
```

2.1.4 VoiceXML-dokumentin toiminnallisuus

2.1.4.1 Alidialogit

Alidialogit (subdialog) ovat VoiceXML:n mekanismi, jolla voidaan hajottaa monimutkaisia dialogeja pienempiin osiin sekä muodostaa uudelleenkäytettäviä komponentteja. Alidialogiin mentäessä päädialogin toiminta keskeytyy ja jatkuu kun se saa alidialogilta paluuarvoja. Kuva 4 havainnollistaa siirtymää päädialogista alidialogiin.



Kuva 4: Alidialogin käyttö VoiceXML sovelluksesta

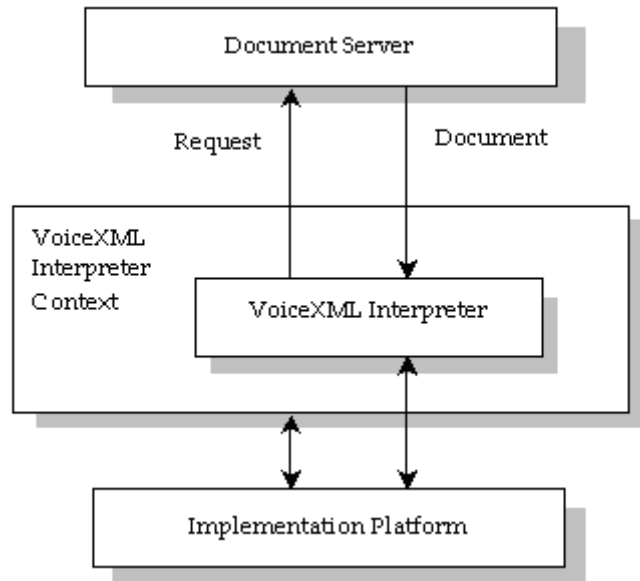
2.1.4.2 Tapahtumat

VoiceXML:n tapahtumat tarjoavat mekanismin, jolla voi käsitellä sellaiset tapahtumat, joita VoiceXML:n lomakkeella ei voida käsitellä.

2.2 Palvelinpuolen ratkaisut ja toteutusarkkitehtuuri

2.2.1 Arkkitehtuurimalli

VoiceXML:n määrittelyn mukainen arkkitehtuurimalli näkyy kuvassa 2.



Kuva 5: VoiceXML toteutuksen arkkitehtuurimalli

Kuvassa olevan mallin osien toiminnallisuus on jaoteltu seuraavanlaisesti: Dokumenttipalvelin (Document Server) huolehtii yleisestä logiikasta, suorittaa tietokantaoperaatioita ja on yhteydessä muihin tietojärjestelmiin. Tämän lisäksi dokumenttipalvelin käsittelee pyyntöjä VoiceXML-tulkilta ja tuottaa vastaukseksi VoiceXML-dokumentteja. VoiceXML-tulkki käsittelee dokumenttipalvelimelta saamansa vastaukset ja toimii niiden logiikan mukaan. VoiceXML-tulkin käyttöympäristö (VoiceXML Interpreter Context) ja VoiceXML-tulkki tarkkailevat käyttäjän syötteitä ja kontrolloivat toteutusalustaa. Toteutusalusta (Implementation Platform) puolestaan muodostaa tapahtumia käyttäjän toiminnan mukaan, joihin VoiceXML-tulkin käyttöympäristö (VoiceXML Interpreter Context) ja VoiceXML-tulkki (VoiceXML Interpreter) sitten reagoivat lähettämällä tarvittaessa pyyntöjä dokumenttipalvelimelle. VoiceXML-tulkin käyttöympäristö ja VoiceXML-tulkki eroavat siten, että VoiceXML tulkin käyttöympäristö hoitaa yleisempiä asioita kuten käyttäjän yhteydenoton havaitsemisen, siihen vastaamisen ja VoiceXML-dokumentin hankinnan. VoiceXML-tulkki huolehtii puolestaan vastauksen jälkeen tapahtuvan dialogin toiminnasta. [VoiceXML, 2004]

2.3 Oheiskielet

VoiceXML kuuluu vuonna 1999 perustetun W3C:n Voice Browser Working Group:in kehittämään Speech Interface Frameworkiin [W3C VBA, 2005]. Speech Interface Frameworkin tarkoituksena on mahdollistaa sellaisten verkkosovellusten teko, joita voidaan käyttää millä tahansa puhelimella puheen ja puhelinten näppäimien kautta. Speech Interface Framework koostuu useasta merkkäusmäärittämisestä, joilla on kullakin kokonaisuudessa oma tehtävänsä. VoiceXML hoitaa Speech Interface Frameworkissa puhedialogin toiminnallisuutta. Tämän lisäksi Framework kattaa puhe-synteesiin liittyvät kielet (SSML [SSML, 2004], PLS [PLS, 2005]), puheentunnistukseen liittyvät kielet (SRGS [SRGS, 2004], SISR [SISR, 2003]), puhelinkeskuksen toiminnallisuutta hoitavan kielen (CCXML [CCXML, 2005]) ja lisäksi pyrkii myös vastaamaan muihin tarpeisiin, jotka liittyvät kyseiseen aihealueeseen. Seuraavissa alakappaleissa näitä kuvaillaan hieman tarkemmin.

2.3.1 SRGS – The Speech Recognition Grammar Specification

SRGS (Speech Recognition Grammar Specification **[SRGS, 2004]**) on kieliopin määrittelyyn tarkoitettu XML-määrittely ja kattaa sekä puheen, että puhelimen näppäimillä lähetettävät DTFM (Dual-tone multifrequency) syötteet. DTFM syötteet ovat tarpeen meluisissa paikoissa ja tilanteissa, jossa puhuminen on kiusallista. SRGS:n avulla voidaan määrittellä sanastoja ja puhemalleja puheentunnistusohjelmaa varten, jotta se pystyisi mahdollisimman hyvin poimimaan puheesta oleellisen. SRGS:llä määritellään yksittäisiä tunnistettavia sanoja sekä niiden järjestys ja puheasu. Uusin SRGS:n määrittelyversio on julkaistu maaliskuussa 2004. **[SRGS, 2004]**

2.3.2 N-Gram specification

N-Gram on ns. stokastinen kielioppimalli (stochastic language model) ja se tarjoaa mahdollisuudet suuren tai täysin avoimen sanaston käyttöön sovelluksessa. Tämä tapahtuu siten, että N-Gram pysyy mallintamaan todennäköisyydet siitä, että joku tietty sana esiintyy joidenkin muiden sanojen jälkeen. Tämä antaa suuntaa siitä mikä seuraava sana todennäköisemmin voisi olla ja tätä kautta helpottaa puheen tunnistamista ja tulkinnan aiheuttamaa tiedonkäsittelykuormaa. Uusin N-Gram määrittelyversio on julkaistu tammikuussa 2001. **[SLM, 2001]**

2.3.3 SSML – The speech synthesis specification

SSML tulee sanoista Speech Synthesis Markup Language **[SSML, 2004]**. SSML on yleisluonteinen mekanismi, jolla merkataan tekstiä puheen esittämistä varten, ja sitä voidaan hyödyntää useissa yhteyksissä. SSML:n avulla kontrolloidaan tarkemmin, kuinka syntetisoitu ääni muodostetaan ja esitetään käyttäjälle. Merkkauksella voidaan vaikuttaa esimerkiksi puheen ääntämiseen, äänen voimakkuuteen, korkeuteen, tempoon, ikään, sukupuoleen jne. Sillä voidaan myös tuoda kuultavaksi digitoituja äänisisältöjä, kuten musiikkia sekä puhetta ja yhdistellä niitä syntetisoidun äänen kanssa.

SSML:ää voidaan käyttää lisäksi minkä tahansa puhesyntetisaatiosovelluksen kanssa, mukaan lukien esimerkiksi sähköisten kirjojen lukusovellukset, sähköpostin lukusovellukset jne. SSML:n ensimmäinen versio vuodelta 2004 on W3C suositus. **[SSML, 2004]**

2.3.4 SISR – Semantic Interpretation for Speech Recognition

Semantic Interpretation Specification **[SISR, 2003]** määrittää puheentunnistuskieliopin, jolla voi kuvata tarkemmin semanttista sisältöä Speech Recognition Grammarissa. SISR-tagien (Semantic Interpretation for Speech Recognition) avulla voidaan liittää SRGS:n yhteyteen ohjeita semanttisen sisällön käsittelyyn. Tämä tapahtuu siten, että SISR-määrittelyn mukainen merkkaukset lisäävät SRGS-tagien sisälle. Uusin Speech Interpretation Specificationin määrittelyluonnos (Working Draft) on julkaistu marraskuussa 2004. **[SISR, 2003]**

2.3.5 PLS – Pronunciation Lexicon Specification

Pronunciation Lexicon **[PLS, 2005]** kuvaa sitä foneettista informaatiota, jota käytetään puheentunnistuksessa ja puhesynteesissä. Pronunciation Lexicon on kehitetty, jotta puhe-sovellusten kehittäjät voisivat toimittaa sovelluksen muille osille täydentävää informaatiota siitä kuinka esimerkiksi erisnimet, paikan nimet ja lyhenteet tulee lausua. Uusin Pronunciation Lexiconin määrittelysuunnitelma on julkaistu Helmikuussa 2005. **[PLS, 2005]**

2.3.6 CCXML – Call Control eXtensible Markup Language

CCXML **[W3C CCR, 2001]** on suunniteltu tarjoamaan puheluiden käsittelytuki dialogipohjaisille järjestelmille, kuten VoiceXML:lle. Puhelujen käsittelyllä tarkoitetaan sitä viestinvälitystä, joka liittyy puhelujen kytkentään, tarkkailuun, siirtoon ja katkaisuun. CCXML kuvaa näihin viestinvälityksiin liittyvien tapahtumien sekä tilojen mukaisen syntaksin sekä joukon puhelun kontrolliin käytettäviä elementtejä. CCXML on varsinaisesti suunniteltu täydentämään VoiceXML:ää, mutta sitä voivat käyttää myös muut dialogipohjaiset systeemit. VoiceXML voi käyttää jotain muutakin puhelun käsittelysystemiä kuin CCXML:ää. Uusin CCXML:n määrittelysuunnitelma on julkaistu Kesäkuussa 2005. **[W3C CCR, 2001]**

2.3.7 SCXML – State Chart XML: State Machine Notation for Control Abstraction

SCXML **[SCXML, 2005]** on ehdokas kontrollikieleksi, jota tullaan mahdollisesti käyttämään ainakin VoiceXML 3.0:ssa ja CCXML 2.0:ssa. SCXML kuvaa sekä kontrolleihin liittyvät tapahtumat, että tilat joihin siirrytään tapahtumien myötä. Uusin SCXML:n suunnitelma on julkaistu heinäkuussa 2005. **[SCXML, 2005]**

2.4 Kilpailijat

2.4.1 SALT

Vuonna 2001 muutamia johtavia ohjelmisto- ja laitteistovalmistajia kuten Cisco, Intel, Microsoft, Comverse, SpeechWorks ja Philips perustivat komitean nimeltä SALT Forum. Tavoitteena oli muodostaa vaihtoehtoinen kieli puhekäyttöliittymien suunnitteluun sekä multimodaalisissa että pelkästään ääneen perustuvissa sovelluksissa. Kieli oli tarkoitettu käytettäväksi HTML:n, XHTML:n ja muiden yleisten kielimäärittelyjen kanssa. Hanke onnistui ja tuloksena syntyi SALT -puhekäyttöliittymän merkkäuskieli.

SALT ja VoiceXML ovat molemmat puhekäyttöliittymien merkkäuskieliä, joskin VoiceXML on enemmän puhelinsovelluksiin painottunut. SALT on puolestaan kevyt lähestymistapa äänen käyttöön missä tahansa selaimessa. Useimmat ensimmäisen sukupolven SALT -sovellukset ovat puhelimen kautta toimivia, samoin kuin useimmat VoiceXML-sovelluksetkin. VoiceXML ja SALT ovat kumpikin sekä puheella että puhelimen DTMF-äänivalinnoilla (Dual-tone multifrequency) ohjattavissa.

Tällä hetkellä SALT:n käyttökohteita ovat muun muassa toimistosovellusten puheohjaus, erilaiset puhelimella käytettävät sovellukset kuten asiakastuen palvelut, uutispalvelut sekä tilaus- ja varausjärjestelmät. SALT-toteutuksia voidaan tehdä myös valmiisiin www-rakenteisiin liittämällä niihin uusia elementtejä. **[Paavilainen, 2004]**

SALT keskittyy pelkästään itse puhekäyttöliittymään, kun VoiceXML:ssä käsitellään ohjelmointiominaisuuksien avulla puhekäyttöliittymän lisäksi myös vuorovaikutuksessa tarvittavaa ja syntyvää dataa. Näin ollen VoiceXML sisältääkin huomattavasti laajemman kirjon erilaisia tageja. Vanhoihin www-kehitystyökaluihin ja -tapoihin totuneet kehittäjät eivät ole ottaneet VoiceXML:ää avosylin vastaan, väittäen sitä työlääksi opetella – vaatiihan sovelluksen tekeminen myös useiden oheiskieltenkin opettelua (esimerkiksi CCXML ja SSML). Erilaisista lähestymistavoista johtuen kielet eivät ole kuitenkaan riidoissa keskenään, vaan pikemminkin täydentävät toisiaan sopien erilaisiin käyttötarpeisiin. **[E-Gram, 2005]**

Vaikka kielten välille on yritetty saada kaksintaistelun makua, on esitetty myös toiveita yleisestä standardista. Kielten lähestyessä toisiaan toivotaan luotavaksi standardi, jonka pohjalta saataisiin yksi selkeä merkkauskieli puheen hyödyntämiseen erilaisissa laitteistoissa ja sovelluksissa. VoiceXML ja SALT käyttävät molemmat W3C:n standardeja. Sekä VoiceXML että SALT suosittelivat SRGS:n ja SSML:n käyttöä kieliopin ja puheen ulostulon määrittelyyn.

SALTin lisäksi ei tällä hetkellä ole muita merkittäviä kilpailijoita.

3. VoiceXML käytännössä

Ääniselaimet soveltuvat erinomaisesti seuraavan sukupolven puhelinyhteyskeskusten toteutukseen. Puhelinyhteyskeskusten luonne muuttuu tämän myötä entistä enemmän puhelukeskuksesta ääniportaalin suuntaan, johon on integroitu myös verkkosisältöjä. Samoja sisältöjä ja palveluita voi käyttää esimerkiksi tavallisen puhelimen ja internetin välityksellä.

Mahdollisia sovellusalueita ovat esimerkiksi erilaiset tilaus-, pankki-, informaatio- ja muut hyötypalvelut, kuten ääniohjatut sähköpostit.

VoiceXML:n kehitys jatkuu ja seuraavassa 3.0 versiossa on luvattu muun muassa parempi yhteensopivuus muiden W3C kielten kanssa. Parannuksia on luvattu tehtävän dialogien ja mediatiedoston hallintaan. Merkittäviä uusia ominaisuuksia ovat myös parannettu modularisaatio, selkeämpi erottelu dialogien ja datan välillä ja asynkroninen tapahtumankäsittely ulkoisille tapahtumille. VoiceXML kehitystyökaluja on runsaasti markkinoilla. Niitä ovat kehittäneet useat merkittävät valmistajat, kuten HP, IBM ja Motorola sekä lukuista joukko pienempiä toimijoita. **[W3C VBA, 2005]**

3.1 Esimerkkejä VoiceXML:n käyttöön otosta

3.1.1 E*TRADE FINANCIAL

Pankki- ja osakepalveluja tarjoava E*TRADE FINANCIAL [E*TRADE, 2005] päätti parantaa tehokkuuttaan ja laajentaa palveluaan lisäämällä puhelinpalveluihinsa normaalin asiakaspalvelun ohkeen VoiceXML-pohjaisen itsepalvelusysteemin. Hankkeessa integroitiin samalla yhtiön puhelin- ja verkkopalvelut sekä kyettiin hyödyntämään myös jo olemassa olevia verkkosisältöjä. Kun palveluiden taustalle saatiin toimintaan yhteinen laitteisto ja liiketoimintalogiikka, voitiin luopua erillisten infrastruktuurien kehittämisestä ja ylläpitämisestä eri asiakasryhmille.

VoiceXML standardin mukainen kehitystyö mahdollisti nopean etenemisen, jossa ei tarvittu uudentyypisiä laitehankintoja tai lisensoituja ohjelmistoja. E*TRADE:n verkkoympäristö oli jo siinä valmis alusta järjestelmän luomiseen. Kun puhelin- ja verkkopalvelut toimivat yhdessä, on uusien puhelinpalveluiden kehittäminen onnistunut yhtä sujuvasti kuin verkkopalveluidenkin.

Uuden puhelinpalvelumuodon ja integraatioprosessin myötä E*TRADE:n palveluiden käyttö on virtaviivaistunut. Esimerkiksi jokaisella E*TRADE:n asiakkaalla on sama salasana sekä puhelin- että verkkopalveluun. Järjestelmä on lisännyt sekä palveluiden käyttömukavuutta että saatavuutta, mutta tuonut myös tuntuvaa taloudellista etua. VoiceXML-pohjaisen järjestelmän on laskettu säästävän vuosittain 30 miljoonaa dollaria. [VoiceXML Forum, 2005]

3.1.2 Spain's Bankinter

Espanjan johtaviin internet pankkeihin ja välittäjiin kuuluva Bankinter suoritti mittavia tutkimuksia, jotka osoittivat, että sen asiakkailla oli olemassa laajalle levinnyt tarve multimodaaliseen käyttöliittymään, joka tekisi pankin mobiilisovellukset käyttökelpoisemmaksi ja käyttäjäystävälliseksi. Bankinter:in mobiilisovelluksilla hoidetaan mm. osakekauppaa ja tilikirjanpitoa. Bankinter:in asiakkaat halusivat, että heillä on mahdollisuus saada joustavasti finanssitietoa ja suorittaa liiketoimia kännykkänsä välityksellä, mutta olivat tyytymättömiä olemassa olevaan naksuttelukäyttöliittymään.

Bankinter ryhtyi toimenpiteisiin kehittämään uutta käyttöliittymää palveluilleen ja päätyi ratkaisuun, jossa puheohjauksella olisi merkittävä osuus uudessa käyttöliittymässä. Yksi Bankinterin kriteereistä uuden sovelluksen ja sen käyttöliittymän alustan suhteen oli, että ne rakennettaisiin yhteensopivaksi vallalle oleviin standardeihin.

Bankinter päätyi ratkaisuun, jossa oli Java:lla rakennettu visuaalinen komponentti ja VoiceXML:llä rakennettu äänikomponentti. Uudistettu käyttöliittymä tarjoaa mm. puhepohjaisen etsintätoiminnon ja tiedonsyötön puhumalla. Puheen tulokset näkyvät myös ruudulla, jotta käyttäjä voi olla varma tietojen oikeellisuudesta. Bankinter on vakuuttanut, että uusi käyttöliittymä kasvattaa heidän asiakastyöväisyyttään sekä lisää pankin itsepalvelutoimintojen määrää.

3.2 Demonstraatio

Animaatio sähköpostin puhelinkäytöstä (Voice-Enabled E-Mail Reader) löytyy osoitteesta: (http://www.voicegenie.com/Phone_Demos.htm?5.0.0.0)

4. Yhteenveto

Ääniohjautuvat ja äänimuotoista palautetta antavat järjestelmät soveltuvat hyvin tiettyihin erikoistarkoituksiin, kuten hankaliin olosuhteisiin tai fyysisiä rajoitteita omaaville käyttäjille. Vaikuttaakin varsin vahvasti siltä, ettei graafinen käyttöliittymä ole korvautumassa valtavirtakäyttäjien keskuudessa.

Suurin syy miksi äänikäyttöliittymät eivät ole yleistyneet nykyistä enemmän on se, että graafisessa käyttöliittymässä on normaalissa käyttötilanteessa huomattavasti enemmän etuja moniin tarkoituksiin. Multimodaaleissa käyttöliittymissä ääni on tärkeä elementti ja tuo etuja pelkkään grafiikkaan nähden. Äänikäyttöliittymät palvelevat tällä hetkellä enimmäkseen erikoiskäyttäjryhmien tarpeita. Tällaisia käyttäjäryhmiä ovat esimerkiksi sokeat ja lukutaidottomat. Äänikäyttöliittymiä käytetään myös tilanteissa, joissa kunnollisen graafisen käyttöliittymän vaatimaa työasemaympäristöä ei ole saatavilla ja käyttäjän täytyy kyetä monimutkaisempaan tiedonvaihtoon järjestelmän kanssa.

VoiceXML on suunniteltu toteuttamaan audiodialogeja, joissa voidaan käyttää synteettistä puhetta, digitalisoitua ääntä, puheentunnista sekä DTFM-syötteitä. Käyttäjän yhtäjaksoista vuorovaikutusta VoiceXML toteutuksen kanssa sanotaan istunnoksi. Instunto koostuu yhdestä tai useammasta VoiceXML-sovelluksesta. VoiceXML sovellukseksi sanotaan joukkoa VoiceXML-dokumentteja, joilla on sama juuridokumentti. VoiceXML-dokumentti koostuu elementeistä, joiden rakenne määräytyy DTD-dokumentin mukaan. Tärkeimmät elementit ovat dialogielementit form ja menu. VoiceXML valikoissa käyttäjälle esitetään ja häneltä kerätään informaatiota. Käyttäjä voi olla vain yhdessä valikossa kerrallaan. Valikoista siirrytään aina seuraavaan niin kauan kunnes seuraavaa ei enää ole, jolloin VoiceXML-sovelluksen suoritus loppuu. Siirtyminen dialogien välillä tapahtuu dialogirakenteen logiikan sekä käyttäjän syötteiden mukaan.

VoiceXML:ää kehitetään osana W3C:n Speech Interface Framework:ia (SIF). SIF:n tavoitteena on mahdollistaa verkkosovellusten käyttö puheen avulla. Puheen syöttöön voi käyttää mitä tahansa puhelinlaitetta. SIF koostuu useasta merkkäusmäärittämisestä, joilla kullakin on oma tehtävänsä kokonaisuudessa. VoiceXML:n osuus SIF:ssä on puhedialogien hoitaminen.

SALT on kehitetty vuonna 2001 vaihtoehtoiseksi toteutuskieleksi VoiceXML:lle. Sitä voidaan käyttää käyttäen HTML:n, XHTML:n ja muiden standardien kanssa puhekäyttöliittymien suunnitteluun erilaisiin ympäristöihin. SALT keskittyy pelkästään itse puhekäyttöliittymään, kun taas laajempi kieli VoiceXML sisältää myös ohjelmointisyntaksin ja datan käsittelymahdollisuuksia. Erilaisista lähestymistavoista johtuen kielet eivät ole riidoissa keskenään, vaan pikemminkin täydentävät toisiaan. VoiceXML ja SALT käyttävät molemmat W3C:n standardeja. SALT on tällä hetkellä ainoa vartenotettava kilpailija VoiceXML:lle.

4.1 Loppuyhteenvedo

Tulevaisuus osoittaa, kuinka suosituksi VoiceXML:ää hyödyntävät sovellukset osoittautuvat. VoiceXML:n takana on aktiivinen kehittäjäyhteisö ja teknologiana se menee koko ajan eteenpäin. Jo nykyisellään VoiceXML on osoittanut soveltuvuutensa perinteisten IVR-palveluiden tuotannossa. VoiceXML:ssä on kuitenkin potentiaalia myös laajempaan käyttöön. Teknologian yleinen kehityskulku on menossa yhä mobiilimpaan suuntaan ja tämä tulee muuttamaan myös ihmisten käyttötottumuksia. Kehittäjäyhteisö odottaa ääniohjattavuuden yleistyvän uusissa käyttöliittymiltään täysin puhepohjaisissa systeemeissa sekä myös multimodaaleissa käyttöliittymissä, joissa ääni toimii perinteisen graafisen käyttöliittymän tukena.

4.2 Tenttikysymys

Tenttikysymys: Mikä VoiceXML on ja mihin sitä käytetään? Kerro lyhyesti minkälaisia syötteitä ja tulosteita VoiceXML sovelluksessa voidaan hyödyntää?

Vastaus: VoiceXML on XML-pohjainen merkintäkieli, jota käytetään sellaisten verkkosovellusten luomiseen, joita on mahdollista käyttää puheen avulla. Puheen syöttöön voidaan käyttää mitä tahansa puhelinlaitetta, tai muuta äänensyöttölaitetta. VoiceXML on suunniteltu toteuttamaan audio-dialogeja, joissa voidaan käyttää synteettistä puhetta, digitalisoitua audiota, tunnistaa puhe- sekä DTMF (Dual Tone Multifrequency) näppäinkomentoja, tallentaa puhuttu syöte sekä käydä keskustelua, jossa käyttäjän puhetta kuunnellaan ja siihen reagoidaan sovelluksen tasolta. Näiden lisäksi VoiceXML- toteutuksiin liittyy perinteisen puhelikeskuksen toimintoihin kuuluvia ominaisuuksia kuten puheluun vastaaminen, sen katkaisu sekä siirto. Nämä toiminnot toteutetaan CCXML-liitännäiskielellä (Call Control eXtensible Markup Language).

Lähteet

[DTSI, 2002]

Dream Tech Software India Inc. 2002. VoiceXML 2.0 Developer's Guide. Blacklick, OH, USA: McGraw-Hill Professional.

[Benyon, 2005]

Benyon D., Turner P., Turner S. 2005. Designing Interactive Systems – People, Activities, Contexts, Technologies. Harlow: Addison Wesley.

[Shneiderman, 2005]

Shneiderman B., Plaisant C. 2005. Designin the User Interface – Strategies for Effective Human-Computer Interaction. New York: Pearson Education, Inc.

[E-Gram, 2005]

E-Gram. VoiceXML vs. SALT: Is that the Right Question? Haettu osoitteesta: (<http://www.netbytel.com/e-gram/egramdetail.asp?ID=7>) [15.11.2005]

[Paavilainen, 2004]

Paavilainen J., Valkonen J., Vepsäläinen J. ITKD60–harjoitustyö: SALT – Speech Application Language Tags, 2004. Haettu osoitteesta: (http://www.ad.jyu.fi/digdoc/ITKD60_2004/Harkkatyot/Esitykset_30112004/Salitti/Salitti.xml) [15.11.2005]

[W3C VBA, 2005]

W3C "Voice Browser" Activity, 2005. Haettu osoitteesta: (<http://www.w3.org/Voice/>) [15.11.2005]

[VoiceXML, 2004]

Voice Extensible Markup Language (VoiceXML) Version 2.0, 2004. Haettu osoitteesta: (<http://www.w3.org/TR/2004/REC-voicexml20-20040316/>) [15.11.2005]

[SRGS, 2004]

Speech Recognition Grammar Specification Version 1.0, 2004. Haettu osoitteesta: (<http://www.w3.org/TR/speech-grammar/>) [15.11.2005]

[SSML, 2004]

Speech Synthesis Markup Language (SSML) Version 1.0, 2004. (<http://www.w3.org/TR/speech-synthesis/>)

[PLS, 2005]

Pronunciation Lexicon Specification (PLS) Version 1.0, 2005. Haettu osoitteesta: (<http://www.w3.org/TR/pronunciation-lexicon/>) [15.11.2005]

[SISR, 2003]

Semantic Interpretation for Speech Recognition Working Draft, 2003. Haettu osoitteesta: (<http://www.w3.org/TR/semantic-interpretation/>) [15.11.2005]

[W3C CCR, 2001]

Call Control Requirements in a Voice Browser Framework Working draft, 2001. Haettu osoitteesta: (<http://www.w3.org/TR/call-control-reqs/>) [15.11.2005]

[SCXML, 2005]

State Chart XML (SCXML): State Machine Notation for Control Abstraction 1.0 Working Draft, 2005. Haettu osoitteesta: (<http://www.w3.org/TR/2005/WD-scxml-20050705/>) [15.11.2005]

[SLM, 2001]

Stochastic Language Models (N–Gram) Specification Working Draft, 2001. Haettu osoitteesta: (<http://www.w3.org/TR/ngram-spec/>) [15.11.2005]

[WDN, 2005]

Wireless Developer Network, An Introduction To VoiceXML. Haettu osoitteesta: (<http://www.wirelessdevnet.com/channels/voice/training/voicexmloverview.html>) [15.11.2005]

[VoiceXML Forum, 2005]

VoiceXML Forum, Customer Success Stories. Haettu osoitteesta: (http://www.voicexml.org/success_stories/) [15.11.2005]

[VoxML, 1999]

VoxML 1.1 Language Reference. Haettu osoitteesta: (<http://www.w3.org/Voice/1999/VoxML.pdf>) [15.11.2005]

[Coverpages, 1999]

Cover Pages: SpeechML. Haettu osoitteesta: (<http://xml.coverpages.org/speechML.html>) [15.11.2005]

[CCXML, 2005]

Voice Browser Call Control: CCXML Version 1.0 – W3C Working Draft 11 January 2005.
Haettu osoitteesta: (<http://www.w3.org/TR/2005/WD-ccxml-20050111/>) [15.11.2005]

[E*TRADE, 2005]

E*TRADE FINANCIAL. Haettu osoitteesta: (<https://us.etrade.com/e/t/home>) [15.11.2005]

[VoiceXML DTD, 2004]

VoiceXML Document Type Definition, 2004. Haettu osoitteesta:
(<http://www.w3.org/TR/voicexml20/vxml.dtd>.) [15.11.2005]