



Εθνικόν και Καποδιστριακόν
Πανεπιστήμιον Αθηνών

Τμήμα Πληροφορικής και Τηλεπικοινωνιών

Φωνητικές Διεπαφές Χρήστη-Τεχνολογίες Φωνής

Ενότητα 1: Εισαγωγή - Εφαρμογές

Γεώργιος Κουρουπέτρογλου

koupe@di.uoa.gr



Σκοπός του μαθήματος

- Η κατανόηση της αρχιτεκτονικής και των συστατικών των Φωνητικών (ή Προφορικών) Διαλογικών Συστημάτων
- Κάλυψη των απαιτούμενων δεξιοτήτων, τεχνικών και εργαλείων για την ανάπτυξη πραγματικών εφαρμογών
- Έμφαση στις πρόσφατες τυποποιήσεις: Πρωτοβουλία voiceWeb του W3C
- Μέρος του Human Computer Interaction (HCI): από τα GUI (Graphical User Interfaces) στα VUI (Voice User Interfaces)

Περιεχόμενα ενότητας

Εισαγωγή στις Φωνητικές Διεπαφές Χρήστη και Εφαρμογές τους

Σενάριο Προφορικού Διαλόγου με Υπολογιστές (1 από 5)

Το ξυπνητήρι χτυπά στις 06.30 π.μ. Η Στέλλα στριφογυρίζει λίγο, ανασηκώνεται και ψιθυρίζει «αναβολή». Δέκα λεπτά αργότερα, το ξυπνητήρι ξαναχτυπά. Η Στέλλα ξυπνάει και λέει «κλείσε». Τεντώνεται, χασμουριέται και σηκώνεται από το κρεβάτι. Γυρίζει το κεφάλι προς το μικρό κουτί δίπλα στο κρεβάτι και λέει: «Πώς είναι ο καιρός σήμερα;». «Στην Αθήνα σήμερα, συννεφιά αρχικά με λίγο ήλιο αργότερα, με θερμοκρασία στους 12 βαθμούς το πρωί με αύξηση στους 16 το απόγευμα», έρχεται η απάντηση από το κουτί. «Αύριο;» «Η πρόβλεψη για την Τρίτη είναι ελαφριά βροχή το πρωί και δυνατή βροχή και ίσως καταιγίδες από τα Δυτικά αργότερα το απόγευμα». «Και στη Βαρκελώνη;» «Η πρόβλεψη του καιρού για τη Βαρκελώνη είναι λιακάδα, με θερμοκρασίες από 16 βαθμούς το πρωί έως 25 το απόγευμα». «Και την Κόστα Μπράβα;» «Συγγνώμη», απαντά το κουτί, «δε γνωρίζω το μέρος που ρωτάτε. Παρακαλώ δοκιμάστε να ρωτήσετε για κάποια πόλη ή χώρα. Για παράδειγμα, ποιες πόλεις γνωρίζετε στην Ισπανία;» «Δεν πειράζει», απαντά η Στέλλα με μια ελαφριά έκφραση ενόχλησης ενώ κατευθύνεται στο μπάνιο.

Σενάριο Προφορικού Διαλόγου με Υπολογιστές (2 από 5)

Έπειτα από ένα ντούζ, η Στέλλα μπαίνει στην κουζίνα, πιάνει μια φέτα ψωμί και το βάζει στην τοστιέρα. «Όχι τόσο ξεροψημένο αυτή τη φορά». Πηγαίνει στο ψυγείο, βγάζει ένα κουτί γάλα και παρατηρεί ότι είναι σχεδόν άδειο. «Μην ξεχάσεις να παραγγείλεις ένα κουτί γάλα», λέει στο ψυγείο. «Αργότερα περιμένεις κάποιους φίλους για ζεστή σοκολάτα, ίσως θα έπρεπε να παραγγείλω δύο κουτιά», απαντά το ψυγείο. «Εντάξει», συμφωνεί η Στέλλα.

Σενάριο Προφορικού Διαλόγου με Υπολογιστές (3 από 5)

Έπειτα από το πρωινό η Στέλλα αρχίζει να σκέφτεται τις δουλειές της ημέρας. Πατά ένα κουμπί στη μικρή συσκευή που βρίσκεται στον καρπό του χεριού της που δείχνει την ώρα. Η συσκευή την καλημερίζει και τη ρωτά τι υπηρεσίες θα ήθελε. «Τι συναντήσεις έχω για σήμερα;», ρωτά η Στέλλα. «Έχεις μια συνάντηση με τον Παύλο στις 10, και μετά μια συνάντηση με την Άννα στις 11», απαντά η συσκευή. «Τηλεφώνησε την Άννα», λέει η Στέλλα, και η συσκευή αρχίζει να τηλεφωνεί. Η Στέλλα συνομιλεί με την Άννα για την αλλαγή της ώρας της συνάντησής τους. Λέει στη συσκευή να ορίσει καινούρια ώρα στο ημερολόγιο και να ελέγξει το email της. Όταν τελειώνει λέει «Ευχαριστώ, αυτά για την ώρα» στη συσκευή στον καρπό της και ετοιμάζεται να φύγει.

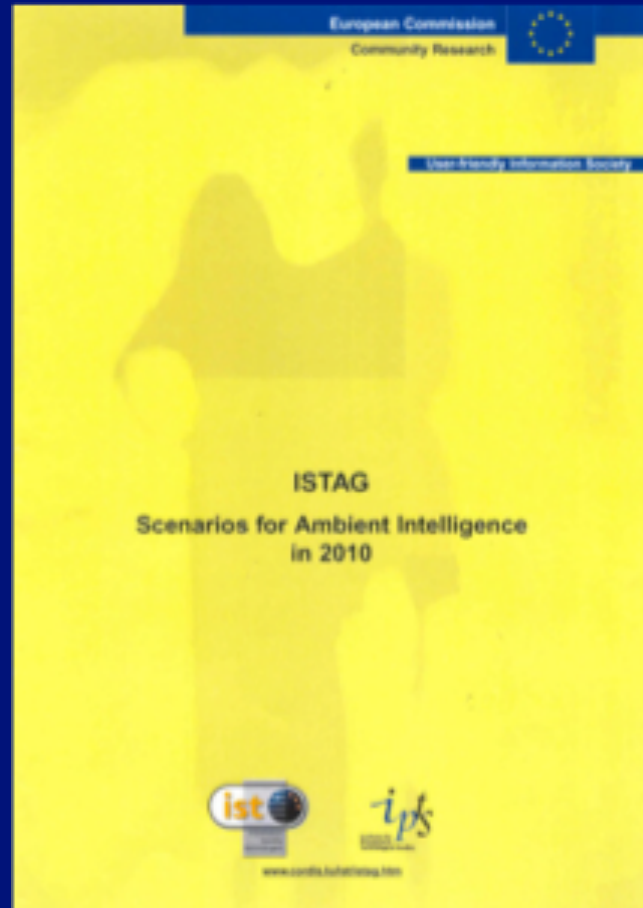
Σενάριο Προφορικού Διαλόγου με Υπολογιστές (4 από 5)

- Maria and P-com: ISTAG (Information Society Technologies Advisory Group) **Scenarios for Ambient Intelligence 2010**, European Commission, 2001

<ftp://ftp.cordis.europa.eu/pub/ist/docs/istagscenarios2010.pdf>

- ΑΣΚΗΣΗ 1.1: Ποιές δραστηριότητες (tasks) που προβλέπει να εκτελεί η Μαρία με φωνητικές εντολές μέσω του P-com είναι σήμερα εφικτές με την υπάρχουσα τεχνολογία που είναι διαθέσιμη στο εμπόριο;

Σενάριο Προφορικού Διαλόγου με Υπολογιστές (5 από 5)



Τρόποι αλληλεπίδρασης Ανθρώπου – Υπολογιστή με Ομιλία (1 από 4)

- Ο χρήστης φορά μικρόφωνο και ακουστικά συνδεδεμένα στον Η/Υ
- Μέσω τηλεφώνου: σύνδεση με απομακρυσμένο Η/Υ (αόρατος στον χρήστη – κυρίως μας ενδιαφέρει εδώ)
- Ενσωματωμένα συστήματα φωνής (π.χ. ηλεκτρικές συσκευές, ρούχα)
- Ενσωματωμένα μικρόφωνα & μεγάφωνα στον περιβάλλοντα χώρο που επικοινωνούν ασύρματα (π.χ. Bluetooth) με Η/Υ (πανταχόθεν & διεισδυτικοί υπολογιστές)

Τρόποι αλληλεπίδρασης Ανθρώπου – Υπολογιστή με Ομιλία (2 από 4)



Τρόποι αλληλεπίδρασης Ανθρώπου – Υπολογιστή με Ομιλία (3 από 4)

Imagine a flat screen display
and wireless microphone.



Τρόποι αλληλεπίδρασης Ανθρώπου – Υπολογιστή με Ομιλία (4 από 4)

This is Mary, one of the talking and listening characters with whom you can interact when using our software on your computer



Τι είναι ένα Σύστημα Προφορικού Διαλόγου (ΣΠΔ);

- Είναι ένα υπολογιστικό σύστημα στο οποίο μπορείς να μιλήσεις για να εκτελέσει κάποια εντολή.
- Τα ΣΠΔ είναι συνήθως δύο ειδών:
 - **Συστήματα παροχής πληροφοριών:** σου παρέχουν πληροφορίες ως απάντηση σε επερώτηση, όπως είναι μια αίτηση για πληροφορίες σχετικά με ένα χρονοδιάγραμμα ή πληροφορίες σχετικά με τον καιρό.
 - **Συστήματα συναλλαγών:** σου επιτρέπουν να αναλάβεις κάποιου είδους συναλλαγή, όπως το να αγοράζεις ή να πουλάς μετοχές ή μία κράτηση θέσης σε αεροπλάνο.

Δύο Εφαρμογές της Τεχνολογίας Αναγνώρισης Ομιλίας

- Στον προσωπικό Η/Υ:
 - εξαρτημένη από τον ομιλητή
 - εκτεταμένο λεξιλόγιο (δεκάδες χιλιάδες λέξεις)
 - εργασίες υπαγόρευσης
- Στην Τηλεφωνία ή Βασισμένα στην Τηλεφωνία:
 - ανεξαρτήτως ομιλητή
 - σχετικά περιορισμένο λεξιλόγιο (εκατοντάδες λέξεις)
 - αμφίδρομες εργασίες

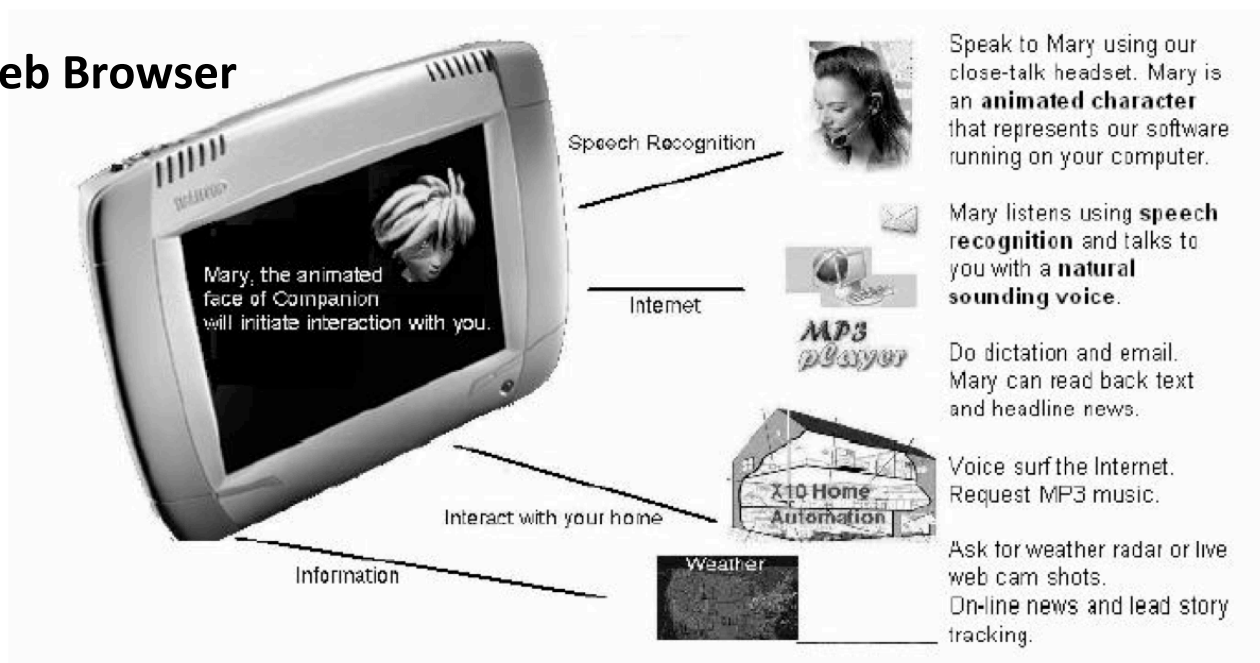
Χρήσεις αναγνώρισης ομιλίας σε προσωπικό Η/Υ (1 από 3)

- Υπαγόρευση
- Ηλεκτρονική αλληλογραφία
- Φωνητικές εντολές
- Πλοήγηση
- Μετάφραση (σε πραγματικό χρόνο) σε άλλη γλώσσα

Χρήσεις αναγνώρισης ομιλίας σε προσωπικό Η/Υ (2 από 3)

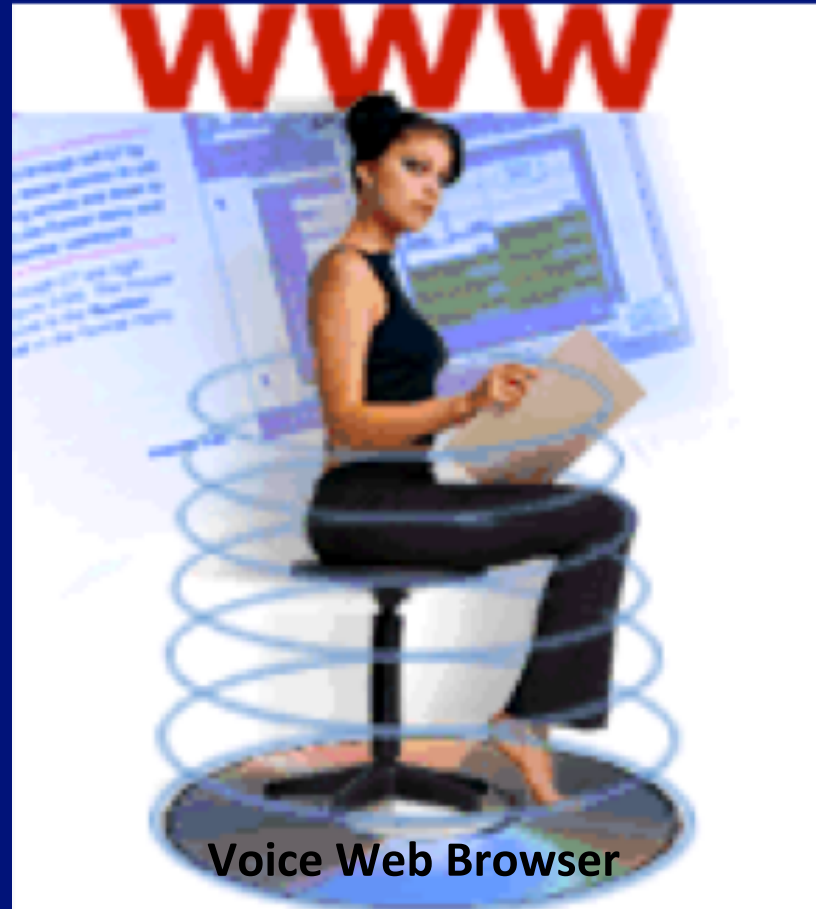
Cutting-edge Technology: Talking Desktop

Voice Web Browser



<http://www.talkingdesktop.com/concept.htm>

Χρήσεις αναγνώρισης ομιλίας σε προσωπικό Η/Υ (3 από 3)

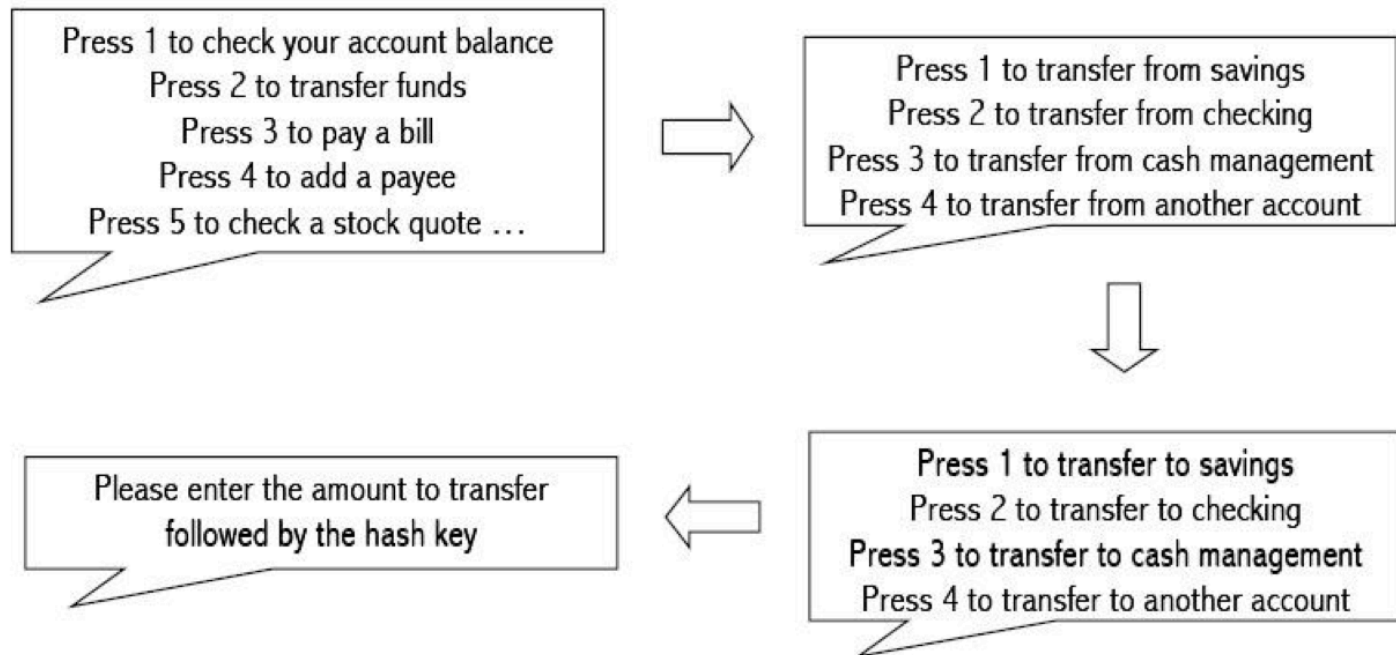


Χρήσεις Τηλεφωνικών Συστημάτων Προφορικού Διαλόγου – Συστημάτων Προφορικού Διαλόγου Βασισμένων στην Τηλεφωνία (1 από 4)

- Απομακρυσμένες τραπεζικές συναλλαγές
- Ταξιδιωτικές κρατήσεις
- Εξέταση Πληροφοριών
- Συναλλαγές μετοχών
- Τηλε-στοιχήματα
- Υπηρεσίες καταλόγου
- Κλήσεις για Taxi
- Παραγγελίες φαγητού

Χρήσεις Τηλεφωνικών Συστημάτων Προφορικού Διαλόγου – Συστημάτων Προφορικού Διαλόγου Βασισμένων στην Τηλεφωνία (2 από 4)

Traditional Interactive Voice Response Systems



Χρήσεις Τηλεφωνικών Συστημάτων Προφορικού Διαλόγου – Συστημάτων Προφορικού Διαλόγου Βασισμένων στην Τηλεφωνία (3 από 4)

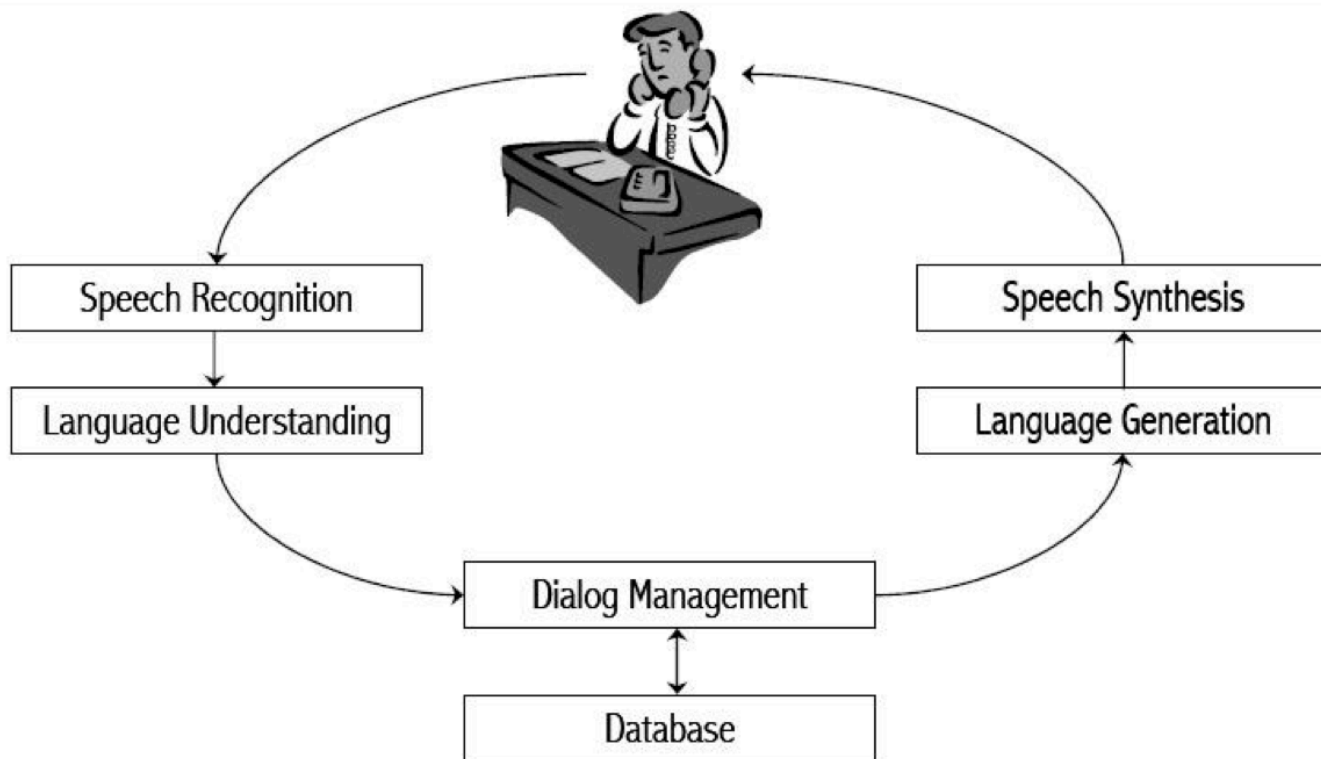
Speech-Enabled Interaction



~25 seconds via speech — as compared with two minutes via touch tone

Χρήσεις Τηλεφωνικών Συστημάτων Προφορικού Διαλόγου – Συστημάτων Προφορικού Διαλόγου Βασισμένων στην Τηλεφωνία (4 από 4)

The Architecture of an SLDS



Τι περιέχει ένα ΣΠΔ;

- **Αναγνώριση ομιλίας:** Αναλύει το εισερχόμενο ηχητικό σήμα ομιλίας ώστε να εξάγει γλωσσικές μονάδες όπως λέξεις ή φωνήματα.
- **Κατανόηση Γλώσσας:** Καθορίζει το νόημα των εισερχόμενων.
- **Διαχείριση διαλόγου:** Ελέγχει τη ροή της συζήτησης, διατηρώντας την ιστορία και το πλαίσιο, κατευθύνοντας την πορεία της, έχοντας πρόσβαση στη βάση δεδομένων, και μορφοποιώντας τις αποκρίσεις.
- **Βάση δεδομένων:** Φυλάσσει τις πληροφορίες που παρέχουν το περιεχόμενο του διαλόγου.
- **Παραγωγή Γλώσσας:** Θέτει τις αποκρίσεις ως λέξεις.
- **Σύνθεση Ομιλίας:** Παράγει το εξαγόμενο ηχητικό σήμα ομιλίας.

Εισαγωγή στον Φωνητικό (Παγκόσμιο) Ιστό (VoiceWeb)

Συνδέοντας το Διαδίκτυο με το Τηλέφωνο

- Ο παγκόσμιος ιστός εξαπλώνεται πέρα από τους προσωπικούς υπολογιστές.
- Τα τηλέφωνα βρίσκονται παντού.
 - Πολύ περισσότερα τηλέφωνα, παρά υπολογιστές.
 - Περισσότερα ασύρματα τηλέφωνα, παρά υπολογιστές, ειδικά στην Ευρώπη.
- Τα ασύρματα τηλέφωνα είναι εξαιρετικά μεταφέρσιμα.
 - Υποστήριξη ενδιαφερουσών υπηρεσιών βάσει τοποθεσίας (location based services).
- Δεν μπορείς να χειριστείς έναν υπολογιστή ενώ οδηγείς.
- Άμεση ενεργοποίηση.

Θέματα

- Αξιοποίηση του Διαδικτύου για ανάπτυξη εφαρμογών και παράδοση υπηρεσιών.
 - Τηλέφωνο αντί για προσωπικό υπολογιστή.
 - VoiceXML αντί για HTML.
 - Ένα φωνητικό πρόγραμμα περιήγησης (Voice Browser) αντί για ένα τυπικό πρόγραμμα περιήγησης.

Παίκτες

- Τελικοί πελάτες
 - Ευκολότερες στη χρήση εφαρμογές
- Πάροχοι Δικτύων
 - Νέο επιχειρησιακό μοντέλο
- Μηχανικοί Ανάπτυξης Εφαρμογών/Περιεχομένου
 - Αξιοποίηση της υποδομής του παγκόσμιου ιστού
 - Μεταφερσιμότητα της τηλεφωνικής υποδομής

Οικονομικά (1 από 2)

- Αύξηση Διείσδυσης διαδικτύου
- Νέες υπηρεσίες
- Νέα ρεύματα εσόδων για επιχειρήσεις
 - Περισσότερες ώρες χρήσης των ασύρματων κυψελοειδών δικτύων
 - Συνδρομητικές Υπηρεσίες

Οικονομικά (2 από 2)

- Παράδοση Υπηρεσιών
 - Φιλοξενία Φωνής vs. Εξοπλισμού
 - Χαμηλό εμπόδια εισόδου
 - Πολλαπλές εφαρμογές μοιράζονται δικτυακούς πόρους
- Ανάπτυξη Υπηρεσιών
 - Μοντέλο ανάπτυξης του παγκόσμιου ιστού
 - Προγραμματιστές τηλεφωνίας vs. Προγραμματιστές παγκόσμιου ιστού

Ομάδες Τυποποίησης

- Φόρουμ για την VoiceXML
<http://www.voicexml.org>



- Ομάδα εργασίας της συνεργασίας W3C για Φωνητικά Προγράμματα Περιήγησης <http://www.w3.org/Voice>

Κοινοπραξία Παγκόσμιου Ιστού (W3C) και φωνή (1 από 2)

- Διεθνή Πρότυπα για τον Παγκόσμιο Ιστό
 - HTML, XML, XHTML κ.α.
- Η δραστηριότητα για το Φωνητικό Πρόγραμμα Περιήγησης (Voice Web Activity)
 - Ξεκίνησε τον Οκτώβριο του 1998
 - Ευρεία ποικιλία ενδιαφερόντων:
 - Φωνητικά Επαυξημένα Γραφικά Προγράμματα Περιήγησης
 - «Φωνητική Περιήγηση» διαδικτυακού περιεχομένου μέσω τηλεφώνου
 - Τηλεφωνικές Υπηρεσίες Φωνής μέσω γλωσσών επισημείωσης συγκεκριμένων πεδίων αναφοράς

Κοινοπραξία Παγκόσμιου Ιστού (W3C) και φωνή (2 από 2)

- Ομάδα εργασίας για Φωνητικά Προγράμματα Περιήγησης
 - Ορίζει πρότυπα για:
 - Επισημείωση διαλόγου
 - Τυποποιήσεις Γραμματικών
 - Επισημείωση Σύνθεσης Ομιλίας
 - Σημασιολογία Φυσικών Γλωσσών

Διεπαφές Φωνής: Ποια Λειτουργικότητα; (1 από 2)

- Βασική Αμφίδρομη Φωνητική Απόκριση (IVR)
 - Computer: “For stock quotes, press 1. For trading, press 2. ...”
 - Human: (presses DTMF “1”)
- Βασική Αμφίδρομη Φωνητική Απόκριση με Ομιλία
 - C: “Say the stock name for a price quote.”
 - H: “Lucent Technologies”

Διεπαφές Φωνής:

Ποια Λειτουργικότητα; (2 από 2)

- Εξελιγμένη Αμφίδρομη Φωνητική Απόκριση με Ομιλία
 - C: “Stock Services, how may I help you?”
 - H: “Uh, what’s Lucent trading at?”
- Αμφίδρομη Φωνητική Απόκριση με “Σχεδόν Φυσική Γλώσσα”
 - C: “How may I help you?”
 - H: “Um, yeah, I’d like to get the current price of Lucent Technologies”
 - C: “Lucent is up two at sixty eight and a half.”
 - H: “OK. I want to buy one hundred shares at market price.”
 - C: “...”

Εφαρμογές Φωνής και Διαδίκτυο (1 από 3)

- **Ανάκτηση Πληροφοριών**
 - Ειδήσεις, καιρός, αθλητικά, κυκλοφοριακή ενημέρωση, χρηματιστήριο.
- **Ηλεκτρονικές Συναλλαγές** (ηλ. εμπόριο, ηλ. τραπεζικές συναλλαγές, e-tailing, κ.α.)
 - Εξυπηρέτηση Πελατών: εντοπισμός αποστολής πακέτων, κατάσταση λογαριασμού, τηλ. κέντρα.
 - Οικονομικά: τραπεζικές συναλλαγές, αγοραπωλησία μετοχών.

Εφαρμογές Φωνής και Διαδίκτυο (2 από 3)

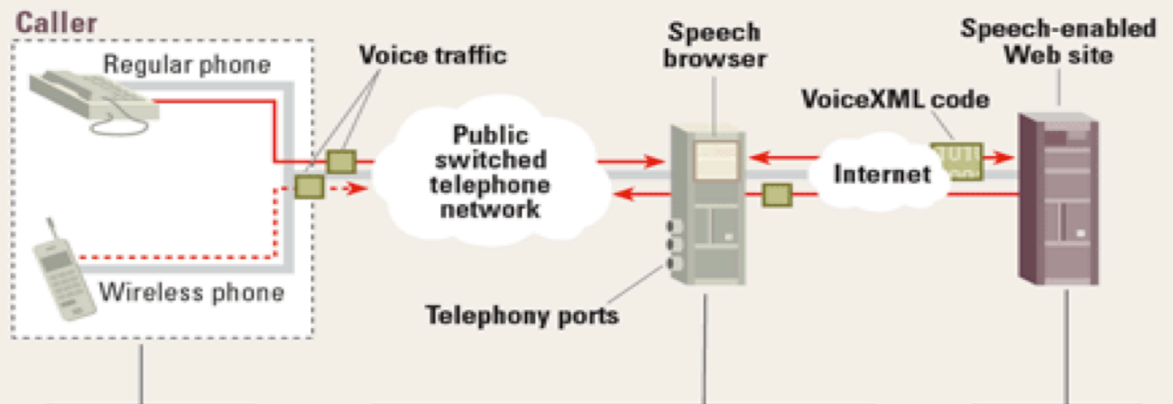
- Τηλεφωνικές Υπηρεσίες
 - Δρομολόγηση κλήσεων, αναγνώστης ηλ. ταχυδρομείου.
 - Προσωπικές, φωνητικά ενεργοποιούμενες, κλήσεις.
 - Υπηρεσίες «Βρείτε-Με» με ένα νούμερο.
- Ταξιδιωτικές Υπηρεσίες
 - Οδηγίες Κατεύθυνσης κατά την οδήγηση, Πληροφορίες Πτήσεων.
- Παιχνίδια και Διασκέδαση
 - Ωροσκόπια, κουίζ, μουσική, ταινίες.

Εφαρμογές Φωνής και Διαδίκτυο (3 από 3)

HOW IT WORKS

VoiceXML

VoiceXML allows application developers to create new speech-based Web content.

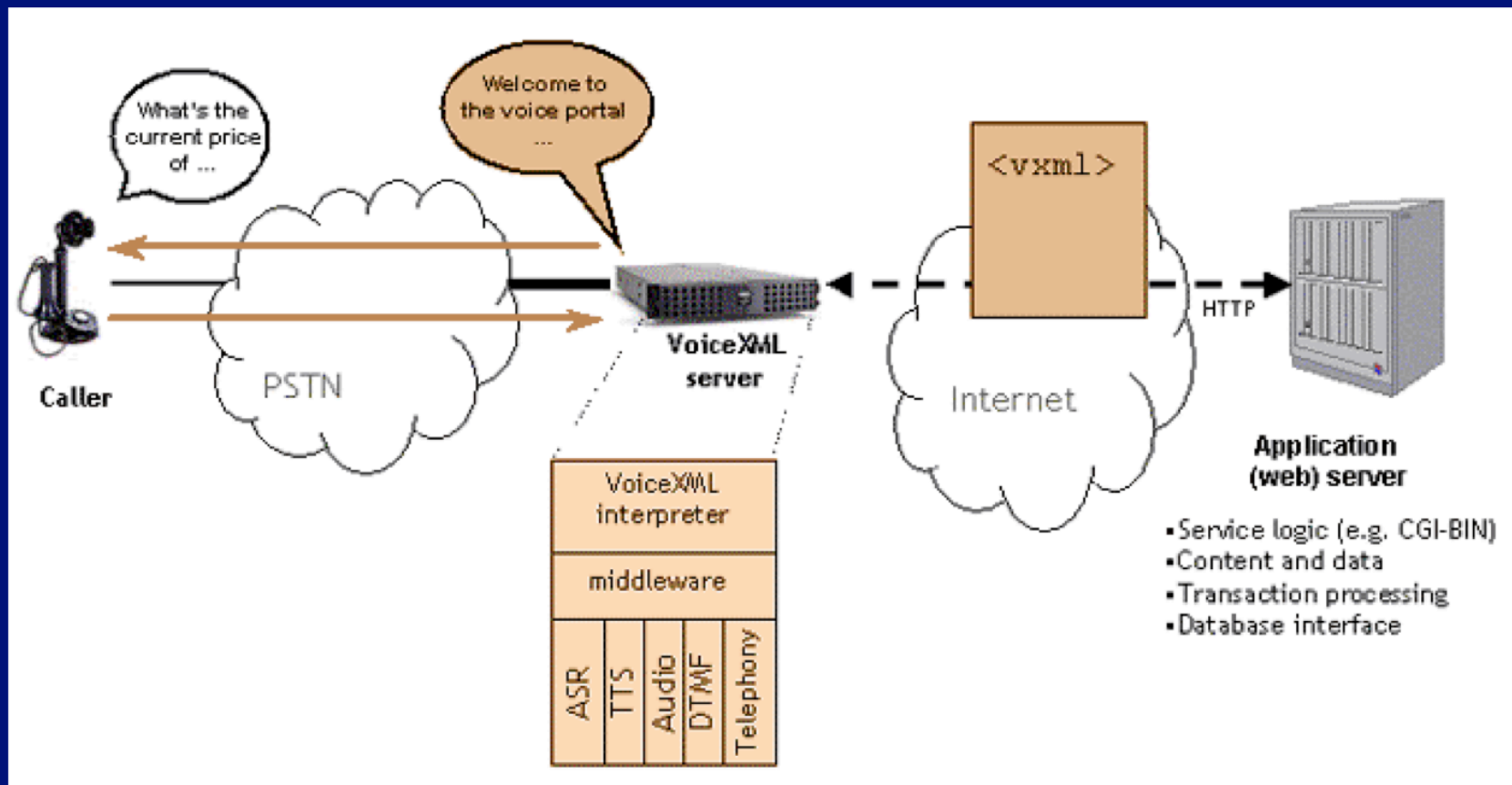


1 A caller can use any voice-enabled device to access a speech-enabled Web site.

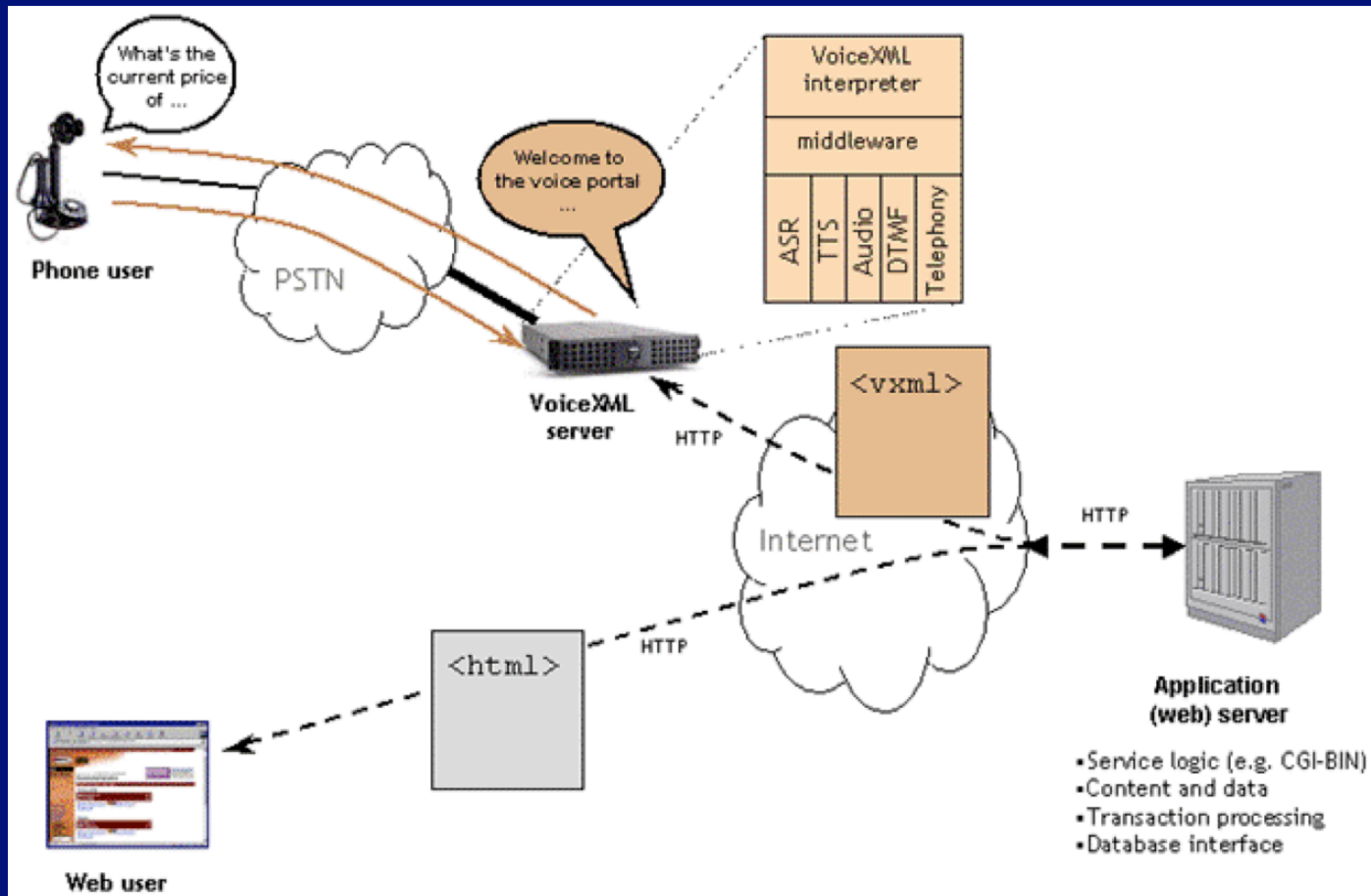
2 A speech browser is the gateway between a voice call and an Internet connection, interpreting VoiceXML code and managing dialog between the caller and VoiceXML content at Web site.

3 A server delivers information back to the caller in audio format.

Αρχιτεκτονική μιας Φωνητικά Ενεργής Διαδικτυακής Εφαρμογής



Σχέση μεταξύ μιας Παραδοσιακής Διαδικτυακής Εφαρμογής και μιας Φωνητικά Ενεργοποιούμενης Διαδικτυακής Εφαρμογής



Κύριες Εταιρείες Τεχνολογιών Φωνής

- **SCANSOFT** www.scansoft.com
- **NUANCE** www.nuance.com
- **IBM** <http://www-01.ibm.com/software/voice/>
- **BEVOCAL** www.bevocal.com
 - <http://www.nuance.com/care/solutions/ondemand/café/>
 - Café VoiceXML development environment
- **VOXPILOT** www.voxpilot.com
 - VocaBase a VoiceXML application server
- **PlumVoice** <http://www.plumvoice.com/developers>
 - free VoiceXML test bed



Κύριες Ελληνικές Εταιρείες

- DIALOGOS
 - www.speech.gr
- OMILIA
 - www.omilia.com
- VOICEWEB
 - www.voiceweb.gr
- e-RHETOR
 - <http://www.e-rhetor.gr/>

Άσκηση 1.2

Διαλογικά Συστήματα Φωνής:

Να παρουσιάσετε σε πίνακα τα προϊόντα και τις υπηρεσίες (με τα κύρια χαρακτηριστικά τους) για όλες τις εταιρείες της προηγούμενης σελίδας.

Άσκηση 1.3

Παρουσιάστε τα Διαλογικά Συστήματα Φωνής (είδη, κύρια χαρακτηριστικά, δυνατότητες) στις περιπτώσεις των Ενσωματωμένων Συστημάτων της Sensory Inc.:

www.sensoryinc.com

(ολοκληρωμένα κυκλώματα και ενσωματωμένο λογισμικό για καταναλωτικά προϊόντα, PDAs, κινητά τηλέφωνα, παιχνίδια, αυτοκίνητα, κιόσκια με verbots)

i) Φωνητική Διαλογική Αλληλεπίδραση

- Χ: Σβήσε τα φώτα (αμφισημία)
- Η: Υπάρχουν περισσότερα από ένα: του καθιστικού ή της κουζίνας; (ερώτηση διασαφήνισης αμφισημίας)
- Χ: Της κουζίνας
- Η: Τα φώτα της κουζίνας είναι τώρα σβηστά
- Η: Με την ευκαιρία, ενεργοποιήθηκε ο αισθητήρας υγρασίας της κουζίνας (πληροφόρηση για μη κρίσιμο γεγονός)

ii) Βελτιωμένη Φωνητική Διαλογική Αλληλεπίδραση (1 από 2)

- Χ: Σβήσε τα φώτα (αμφισημία)
- Η: Υπάρχουν περισσότερα από ένα: του καθιστικού ή της κουζίνας; (ερώτηση διασαφήνισης αμφισημίας)
- (Ο αισθητήρας καπνού στο υπνοδωμάτιο πλησιάζει το επίπεδο συναγερμού)
- Η: ΠΡΟΣΟΧΗ! Έχει ενεργοποιηθεί ο αισθητήρας καπνού στο υπνοδωμάτιο (κρίσιμο γεγονός- διακοπή τρέχοντος διαλόγου)
- Ο χρήστης τρέχει πάνω και βρίσκει τον νεαρό γιο του να καπνίζει στο υπνοδωμάτιο

ii) Βελτιωμένη Φωνητική Διαλογική Αλληλεπίδραση (2 από 2)

- Χ: ΟΚ. Λάθος συναγερμός.
- *Το σύστημα κλείνει τον υποδιάλογο*
- Η: Και τώρα, υπάρχουν περισσότερα από δύο φώτα αναμμένα.
- Χ: Σβήσε της κουζίνας
- Η: Τα φώτα της κουζίνας είναι τώρα σβηστά
- Η: Με την ευκαιρία, ενεργοποιήθηκε ο αισθητήρας υγρασίας της κουζίνας (πληροφόρηση για μη κρίσιμο γεγονός)

1) Έλεγχος Συσκευών με Φωνή (1 από 2)

- Λειτουργία Command-and-control
- Από εφαρμογές λογισμικού – οικιακές ηλεκτρικές συσκευές
- Κατάλληλα και για περιπτώσεις «απασχολημένων χεριών» π.χ. Αυτοκίνητο, χρήση ποντικιού
- Χρήσιμη για άτομα με αναπηρίες που δεν μπορούν να χειριστούν τις συσκευές ή το λογισμικό
- Πολλές φορές με μικρό λεξιλόγιο (10-50 λέξεις), άλλες φορές (π.χ. Πιλοτήριο αεροπλάνου) ~5000 λέξεις
- Σε ενσωματωμένα συστήματα 50-100 φράσεις εξαρτώμενες από τον ομιλητή και 10-20 σε φράσεις ανεξάρτητες του χρήστη

1) Έλεγχος Συσκευών με Φωνή (2 από 2)

- Συνήθως άμεση αντιστοίχιση στη σημασιολογία των λέξεων ή φράσεων
- Πιο σύνθετες φράσεις, όπως «άνοιξε την πόρτα του γκαράζ στις 7:00»
- Αλλά: π.χ. «κάθε Σάββατο και Κυριακή άνοιγε τα φώτα του καθιστικού στις 18:00 για 4 ώρες» πολλές εναλλακτικές λέξεις, πολλοί πιθανοί συνδυασμοί ενεργειών > ανάγκη χρήσης γραμματικής η οποία προσδιορίζει με πιο αφηρημένο τρόπο τις επιτρεπόμενες λέξεις, φράσεις και τους συνδυασμούς τους.
- Διαφορετικοί τρόποι έκφρασης για το ίδιο πράγμα: «άναψε τα φώτα στις 7», «στις 7 άναψε τα φώτα», «τα φώτα άναψέ τα στις 7»

2) Εισαγωγή Δεδομένων με Φωνή

- Σε εφαρμογή λογισμικού (π.χ. Συμπλήρωση πεδίων φόρμας, τιμολογίων, αλλαγή ραντεβού)
- Μικρός αριθμός συνηθισμένων λέξεων
- Ίσως μεγάλος αριθμός λέξεων εξαρτώμενων από την εφαρμογή (π.χ. Ένα σύστημα αναφορών τροχαίων δυστυχημάτων απαιτεί όλα τα ονόματα των δρόμων και των τοποθεσιών μιας πόλης)

3) Λήψη Πληροφοριών με Φωνή

- Αρκετές εφαρμογές σε λειτουργία
- Χρήση σχεδόν φυσικής ομιλίας
- MIT: σύστημα πληροφοριών καιρού Jupiter + κρατήσεων θέσεων Pegasus (4000 πτήσεις) + Voyager για καθοδήγηση πορείας και κυκλοφορίας στην Boston (1977-1999: 30.000 κλήσεις, 80% επιτυχής αναγνώριση για νέους χρήστες)
- Μεγάλο μέγεθος λεξιλογίου

4) Φωνητική Υπαγόρευση

- Σε πραγματικό χρόνο ή από προεκφωνημένη ομιλία
- Επιτρέπουν συνδυασμό φωνητικών εντολών για μορφοποίηση του κειμένου
- Δεν υπάρχει διάλογος με τον χρήστη
- Μπορεί όμως να ενσωματωθεί σε εφαρμογές (π.χ. Mailer)
- Πολύ μεγάλο λεξικό ή και εξειδικευμένο: μικρό λεξικό > καλύτερη ακουστική αναγνώριση (κυρίως σε ακουστικά ίδιες λέξεις), πολύ μεγάλο λεξικό > μικρότερα λάθη εισαγωγής λανθασμένων λέξεων
- Μεγάλο πρόβλημα οι άγνωστες λέξεις που αντικαθίστανται από κάποια του υπάρχοντος λεξικού

5) Φωνητικός Προσωπικός Βοηθός Speech (Voice) enabled Virtual Assistant

- Χειρίζεται έναν αριθμό αυτόματων λειτουργιών: βρίσκει και καλεί τηλεφωνικούς αριθμούς, λαμβάνει e-mails, συμβουλεύεται και ενημερώνει ημερολόγιο συναντήσεων ή γεγονότων
- Π.χ. «κάλεσε τον Γιάννη στο γραφείο», «Ραντεβού της επόμενης Δευτέρας από τις 5 ως τις 7 το απόγευμα», «Διάβασέ μου το e-mail από τον Πρόεδρο»
- Π.χ. VIALTO, VoiceGenie VoiceXML Gateway

iii) Φωνητικά Συστήματα Προηγμένου Διαλόγου (1 από 4)

- Περιπτώσεις που ο χρήστης εκτελεί πολλές εξαρτώμενες ενέργειες
 - π.χ. Πληροφορίες και κρατήσεις πτήσεων, ενοικίαση αυτοκινήτου, κράτηση ξενοδοχείου
- Επιτρέπουν τον χρήστη να μετακινείται μεταξύ των ενεργειών και το σύστημα θυμάται τι έχει συζητηθεί (π.χ. Να αποφασίσει για ενοικίαση αυτοκινήτου εξαρτώμενη από την τοποθεσία του ξενοδοχείου)

iii) Φωνητικά Συστήματα Προηγμένου Διαλόγου (2 από 4)

- Προσπαθεί να αντιμετωπίζει τα σφάλματα και τις παρανοήσεις
 - Παράδειγμα:
 - Χ: Που υπάρχει λιακάδα στην Καραϊβική;
 - Υ: Συγγνώμη, δεν είμαι σίγουρος τι είπες
 - Χ: Που υπάρχει λιακάδα στη Καραϊβική;
 - Υ: Συγγνώμη, μπορεί να μη γνωρίζω τη πόλη που ενδιαφέρεσαι. Προσπάθησε να ρωτήσεις για την χώρα της ή την ήπειρο.

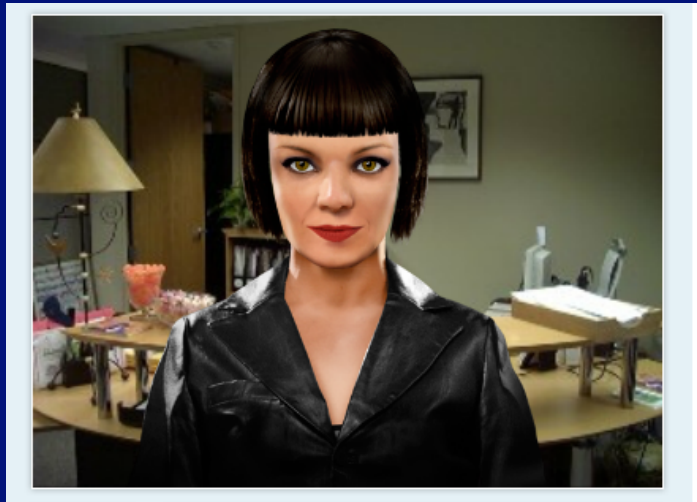
iii) Φωνητικά Συστήματα Προηγμένου Διαλόγου (3 από 4)

- Προσθήκη «λογικής κρίσης» (reasoning) στο σύστημα
 - Παράδειγμα:
 - X: Χρειάζεται να πάμε την ασθενή στο Ιπποκράτειο στην Αθήνα (υπάρχει αντίστοιχο και στη Θεσ/νίκη)
 - Y: OK
 - X: Τι οχήματα είναι διαθέσιμα;
 - Y: Υπάρχουν ασθενοφόρα στην Καρδίτσα και στον Βόλο
 - X: Προτιμώ αυτό του Βόλου
 - Y: Γνωρίζεις ότι ο νότιος δρόμος Βόλου-Εθνικής είναι μπλοκαρισμένος λόγω έργων;

iii) Φωνητικά Συστήματα Προηγμένου Διαλόγου (4 από 4)

- RAMONA: φωτορεαλιστική προσωποποίηση πράκτορα συζήτησης

<http://www.kurzweilai.net/Ramona4.2/ramona.html>



TRY

- ... changing Ramona's Personality
- ... changing Ramona's Accent
- ... asking her about well-know people, books, events
- ... moving your cursor around in her window
- ... earlier versions of Ramona

Maske-bot: A talking video humanoid robot

- November 8, 2011
- <http://www.youtube.com/watch?v=oFp1hpH25oI>
- <http://www.kurzweilai.net/mask-bot-a-talking-video-humanoid-robot>



European Network of Excellence in Human Language Technologies (1 από 2)

- <http://www.elsnet.org/>
- Roadmap: <http://elsnet.dfki.de/>

European Network of Excellence in Human Language Technologies (2 από 2)



Η πρόοδος των Τεχνολογιών Φωνής

Είναι αποτέλεσμα ενός αριθμού παραγόντων:

- Τεχνολογικοί πρόοδοι στο hardware των υπολογιστών
- Ανάπτυξη ειδικού λογισμικού και αλγορίθμων
- Υποδομών για το VoiceWeb
- Εμπορικής Ώθησης

Μέρη Μαθήματος (1 από 2)

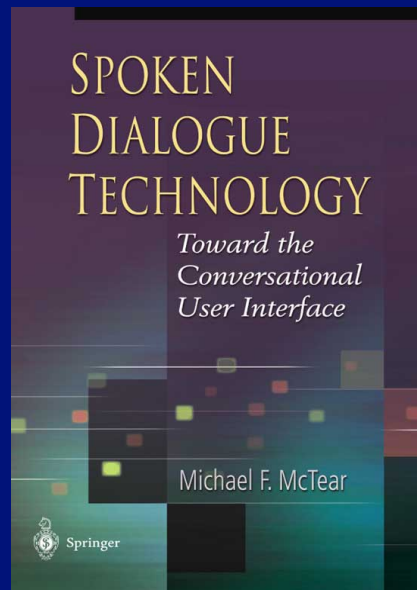
- Εφαρμογές Διαλογικών Εφαρμογών Ομιλίας
- Κατανόηση Διαλόγου (Χαρακτηριστικά & Μοντελοποίηση)
- Παραγωγή Γλώσσας
- Μετατροπή Κειμένου σε Συνθετική Ομιλία
- Αναγνώριση Ομιλίας
- Κατανόηση Γλώσσας

Μέρη Μαθήματος (2 από 2)

- Έλεγχος Διαλόγου
- Ανάπτυξη Διαλογικών Εφαρμογών – Κύκλος ζωής
- Γρήγορη Προτυποποίηση - CSLU toolkit (Center for Spoken Language Understanding – Oregon University)
- Διεθνείς Τυποποιήσεις & ο VoiceWeb
- VoiceXML
- Πολυτροπικές (multimodal) εφαρμογές με Ομιλία:
 - SALT (Speech Application Language Tags)
 - MS .NET Speech SDK (Software Development Kit)

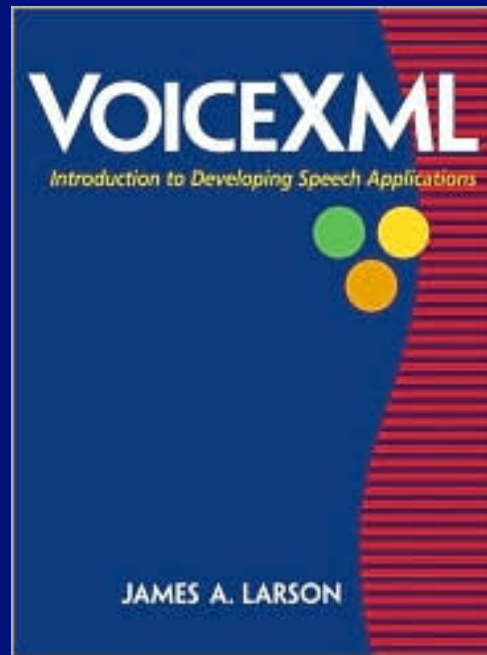
Βιβλία Αναφοράς (1 από 3)

- M. McTear “Spoken Dialogue Technology: Toward the Conversational User Interface” Springer, 2004



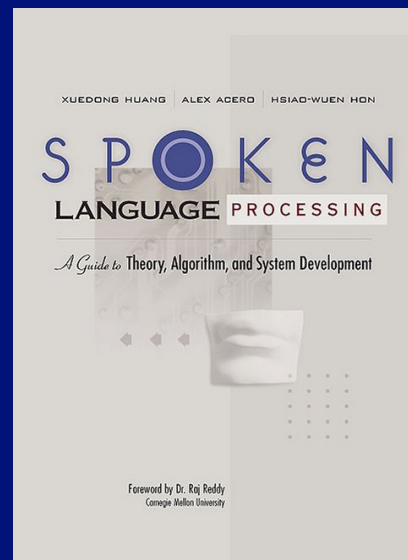
Βιβλία Αναφοράς (2 από 3)

- J. Larson “VoiceXML: Introduction to Developing Speech Applications”, Prentice Hall 2003



Βιβλία Αναφοράς (3 από 3)

- X. Huang, A. Acero, H.-W. Hon “Spoken Language Processing: Theory, Algorithm and System Development” Prentice Hall 2001



Ιστοσελίδα Μαθήματος

<http://eclass.di.uoa.gr>

Απαιτείται εγγραφή

Βαθμός Μαθήματος

- Γραπτή εξέταση: 70%
- Ασκήσεις: 30% (παραδίδονται σε έντυπη μορφή στη θυρίδα του διδάσκοντα και σε ηλεκτρονική έκδοση στην πλατφόρμα *eclass*)

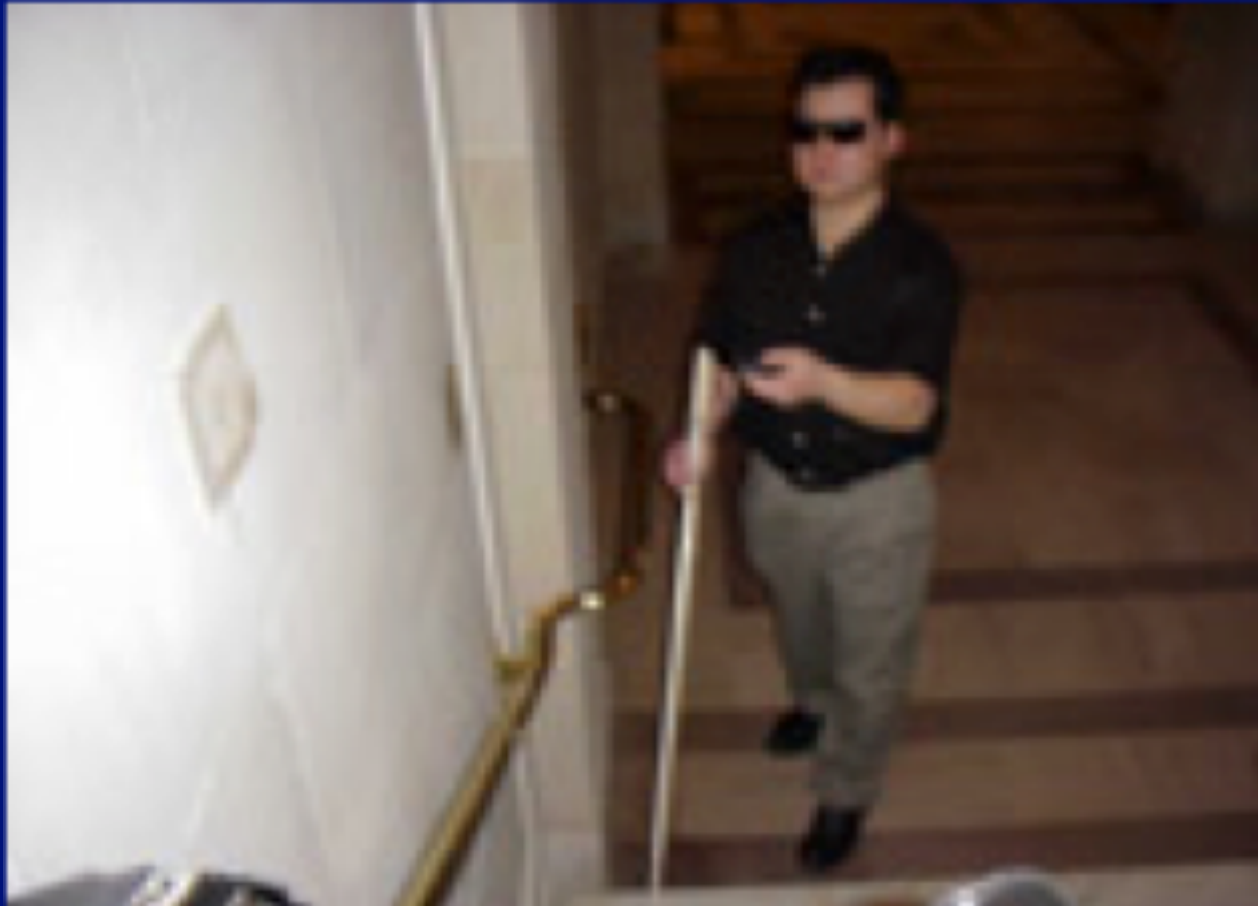
Προσανατολισμός και καθοδήγηση στον δομημένο χώρο – Πληροφόρηση σημάνσεων (1 από 24)

Ερευνητικό Έργο ΜΝΗΣΙΚΛΗΣ:

Προηγμένες Καθολικές Υπηρεσίες Θέσης σε
Εσωτερικούς Χώρους

<http://speech.di.uoa.gr/mnisiklis/>

Προσανατολισμός και καθοδήγηση στον δομημένο χώρο – Πληροφόρηση σημάνσεων (2 από 24)



Προσανατολισμός και καθοδήγηση στον δομημένο χώρο – Πληροφόρηση σημάνσεων (3 από 24)



Προσανατολισμός και καθοδήγηση στον δομημένο χώρο – Πληροφόρηση σημάνσεων (4 από 24)



Προσανατολισμός και καθοδήγηση στον δομημένο χώρο – Πληροφόρηση σημάνσεων (5 από 24)



Προσανατολισμός και καθοδήγηση στον δομημένο χώρο – Πληροφόρηση σημάνσεων (6 από 24)

Ερευνητικό Έργο ΟΜΗΡΟΣ:

Μεθοδολογική Προσέγγιση για τη Σχεδίαση
Ακουστικής-Απτικής Αλληλεπίδρασης στη Μη-
Οπτική Διεπαφή Χρήστη (user interface) με Έμφαση
στην Προσβασιμότητα Ατόμων με Απώλεια Όρασης

<http://speech.di.uoa.gr/homer/>

Προσανατολισμός και καθοδήγηση στον δομημένο χώρο – Πληροφόρηση σημάνσεων (7 από 24)



Προσανατολισμός και καθοδήγηση στον δομημένο χώρο – Πληροφόρηση σημάνσεων (8 από 24)



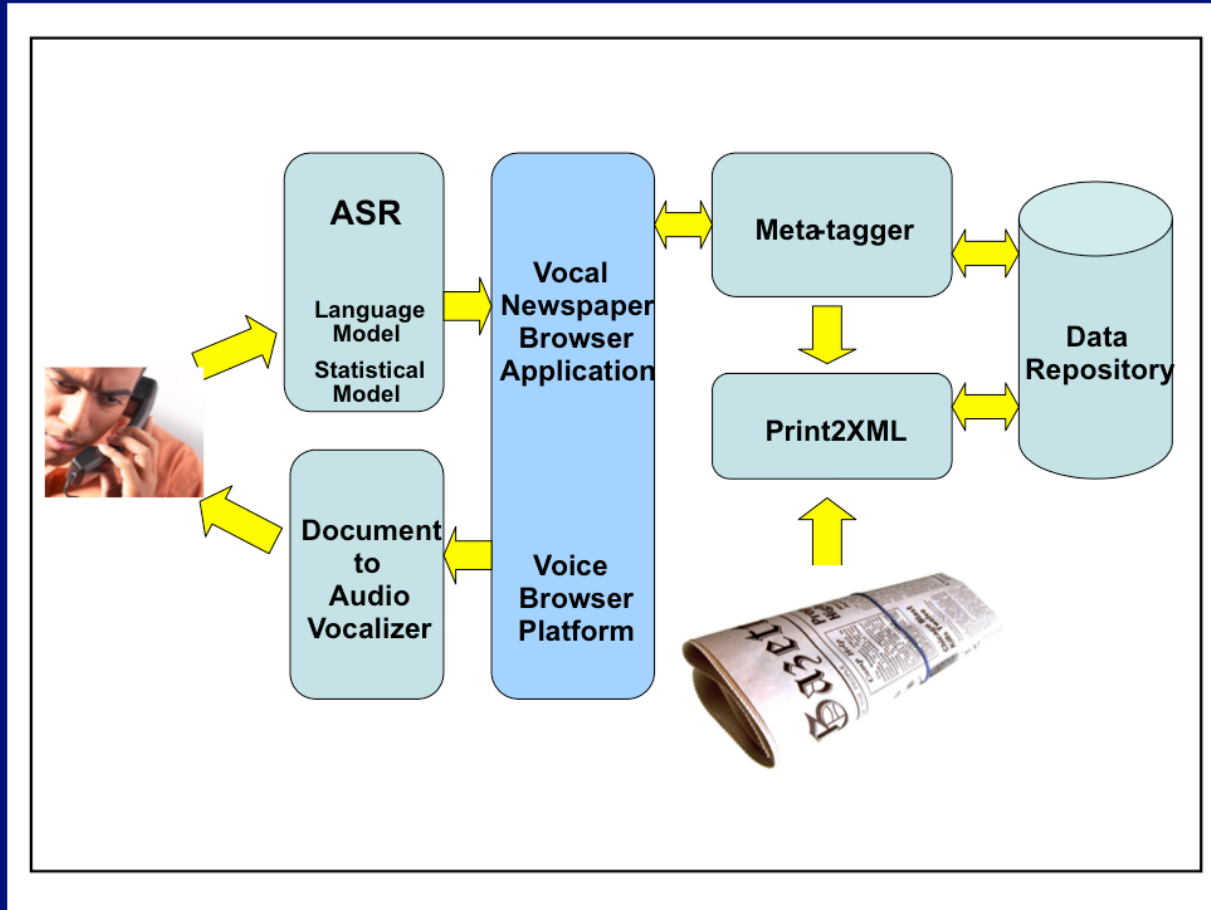
Προσανατολισμός και καθοδήγηση στον δομημένο χώρο – Πληροφόρηση σημάνσεων (9 από 24)

Ερευνητικό έργο ΡΗΤΩΡ:

Συστηματική προσέγγιση διεπαφής φωνητικού
διαλόγου με ακουστικοποίηση της
μεταπληροφορίας και οπτικής δομής εγγράφων με
σκοπό την καθολική πρόσβαση σε έντυπο
περιεχόμενο

<http://speech.di.uoa.gr/rhetor/>

Προσανατολισμός και καθοδήγηση στον δομημένο χώρο – Πληροφόρηση σημάτων (10 από 24)



Προσανατολισμός και καθοδήγηση στον δομημένο χώρο – Πληροφόρηση σημάνσεων (11 από 24)

ΔΗΜΟΣΘΕΝΗΣ:

Σύστημα μετατροπής κειμένου σε συνθετική ομιλία
διατίθεται ελεύθερα: <http://demosthenes.di.uoa.gr/>

- Τεχνολογία συνθετικής φωνής
 - Φυσική ομιλία
 - Προσωδικά χαρακτηριστικά
- Χωρίς προ-ηχογράφηση
 - Εκφωνεί οποιοδήποτε κείμενο
 - Παραγωγή ομιλίας σε πραγματικό χρόνο
- Ταυτόχρονη υποστήριξη πολλών γλωσσών
 - Multilingual & polyglot

Προσανατολισμός και καθοδήγηση στον δομημένο χώρο – Πληροφόρηση σημάνσεων (12 από 24)

Ο ΔΗΜΟΣΘΕΝΗΣ είναι ένα πολυγλωσσικό σύστημα μετατροπής κειμένου σε συνθετική ομιλία, που αναπτύχθηκε στο Τμήμα Πληροφορικής του Πανεπιστημίου Αθηνών. Supports the English, but also μιλά εξίσου καλά και ελληνικά.



Προσανατολισμός και καθοδήγηση στον δομημένο χώρο – Πληροφόρηση σημάνσεων (13 από 24)

Ερευνητικό έργο M-PIRO

<http://www.ltg.ed.ac.uk/mpiro/>



Προσανατολισμός και καθοδήγηση στον δομημένο χώρο – Πληροφόρηση σημάνσεων (14 από 24)



Figure 2: The 3D collection of the M-PIRO virtual reality prototype: spoken descriptions are generated for each selected object.

Προσανατολισμός και καθοδήγηση στον δομημένο χώρο – Πληροφόρηση σημάνσεων (15 από 24)

Ερευνητικό έργο ΑΙΝΕΙΑΣ

Ανάπτυξη Ευέλικτων Συστημάτων Εναλλακτικής και
Επαυξητικής Διαπροσωπικής Επικοινωνίας

<http://speech.di.uoa.gr/aeneas/>



Προσανατολισμός και καθοδήγηση στον δομημένο χώρο – Πληροφόρηση σημάνσεων (16 από 24)



Προσανατολισμός και καθοδήγηση στον δομημένο χώρο – Πληροφόρηση σημάνσεων (17 από 24)



Προσανατολισμός και καθοδήγηση στον δομημένο χώρο – Πληροφόρηση σημάτων (18 από 24)



Προσανατολισμός και καθοδήγηση στον δομημένο χώρο – Πληροφόρηση σημάτων (19 από 24)



Προσανατολισμός και καθοδήγηση στον δομημένο χώρο – Πληροφόρηση σημάνσεων (20 από 24)

Έργο LibAccessATHENA

Υποστήριξη ΑΜΕΑ χρηστών βιβλιοθηκών ΕΚΠΑ

<http://speech.di.uoa.gr/libaccess/>

Προσανατολισμός και καθοδήγηση στον δομημένο χώρο – Πληροφόρηση σημάνσεων (21 από 24)



Προσανατολισμός και καθοδήγηση στον δομημένο χώρο – Πληροφόρηση σημάνσεων (22 από 24)

<http://e-bility.gr>



Προσανατολισμός και καθοδήγηση στον δομημένο χώρο – Πληροφόρηση σημάνσεων (23 από 24)

Συλλογή Δωρεάν Λογισμικού ΑΘΗΝΑ

Στοχεύει να ενημερώσει αλλά και να παρέχει στα
Άτομα με Αναπηρία και τους επαγγελματίες λύσεις
Υποστηρικτικών Τεχνολογιών Πληροφορικής που
δεν έχουν κόστος (Open Source ή Freeware)

<http://access.uoa.gr/ATHENA/>

Προσανατολισμός και καθοδήγηση στον δομημένο χώρο – Πληροφόρηση σημάτων (24 από 24)

Συλλογή Δωρεάν Λογισμικού ΑΘΗΝΑ

Στοχεύει να ενημερώσει αλλά και να παρέχει στα
Άτομα με Αναπηρία και τους επαγγελματίες λύσεις
Υποστηρικτικών Τεχνολογιών Πληροφορικής που
δεν έχουν κόστος (Open Source ή Freeware)

<http://access.uoa.gr/ATHENA/>

Τέλος Ενότητας

Εισαγωγή - Εφαρμογές

Χρηματοδότηση

- Το παρόν εκπαιδευτικό υλικό έχει αναπτυχθεί στο πλαίσιο του εκπαιδευτικού έργου του διδάσκοντα.
- Το έργο «**Ανοικτά Ακαδημαϊκά Μαθήματα στο Πανεπιστήμιο Αθηνών**» έχει χρηματοδοτήσει μόνο την αναδιαμόρφωση του εκπαιδευτικού υλικού.
- Το έργο υλοποιείται στο πλαίσιο του Επιχειρησιακού Προγράμματος «**Εκπαίδευση και Δια Βίου Μάθηση**» και συγχρηματοδοτείται από την Ευρωπαϊκή Ένωση (Ευρωπαϊκό Κοινωνικό Ταμείο) και από εθνικούς πόρους.



Σημειώματα

Σημείωμα Ιστορικού Εκδόσεων Έργου

Το παρόν έργο αποτελεί την έκδοση 1.0.

Σημείωμα Αναφοράς

Copyright Εθνικών και Καποδιστριακών Πανεπιστημίων Αθηνών 2015, Γεώργιος Κουρουπέτρογλου 2015. Γεώργιος Κουρουπέτρογλου. «Φωνητικές Διεπαφές Χρήστη-Τεχνολογίες Φωνής. Εισαγωγή-Εφαρμογές». Έκδοση: 1.0. Αθήνα 2015. Διαθέσιμο από τη δικτυακή διεύθυνση: <http://opencourses.uoa.gr/courses/DI37/>.

Σημείωμα Αδειοδότησης

Το παρόν υλικό διατίθεται με τους όρους της άδειας χρήσης Creative Commons Αναφορά, Μη Εμπορική Χρήση Παρόμοια Διανομή 4.0 [1] ή μεταγενέστερη, Διεθνής Έκδοση. Εξαιρούνται τα αυτοτελή έργα τρίτων π.χ. φωτογραφίες, διαγράμματα κ.λ.π., τα οποία εμπεριέχονται σε αυτό και τα οποία αναφέρονται μαζί με τους όρους χρήσης τους στο «Σημείωμα Χρήσης Έργων Τρίτων».



[1] <http://creativecommons.org/licenses/by-nc-sa/4.0/>

Ως **Μη Εμπορική** ορίζεται η χρήση:

- που δεν περιλαμβάνει άμεσο ή έμμεσο οικονομικό όφελος από την χρήση του έργου, για το διανομέα του έργου και αδειοδόχο
- που δεν περιλαμβάνει οικονομική συναλλαγή ως προϋπόθεση για τη χρήση ή πρόσβαση στο έργο
- που δεν προσπορίζει στο διανομέα του έργου και αδειοδόχο έμμεσο οικονομικό όφελος (π.χ. διαφημίσεις) από την προβολή του έργου σε διαδικτυακό τόπο

Ο δικαιούχος μπορεί να παρέχει στον αδειοδόχο ξεχωριστή άδεια να χρησιμοποιεί το έργο για εμπορική χρήση, εφόσον αυτό του ζητηθεί.

Διατήρηση Σημειωμάτων

Οποιαδήποτε αναπαραγωγή ή διασκευή του υλικού θα πρέπει να συμπεριλαμβάνει:

- το Σημείωμα Αναφοράς
- το Σημείωμα Αδειοδότησης
- τη δήλωση Διατήρησης Σημειωμάτων
- το Σημείωμα Χρήσης Έργων Τρίτων (εφόσον υπάρχει)

μαζί με τους συνοδευόμενους υπερσυνδέσμους.

Σημείωμα Χρήσης Έργων Τρίτων (1 από 2)

Οι φωτογραφίες που περιέχονται στην παρουσίαση αποτελούν πνευματική ιδιοκτησία τρίτων. Απαγορεύεται η αναπαραγωγή, αναδημοσίευση και διάθεσή τους στο κοινό με οποιονδήποτε τρόπο χωρίς τη λήψη άδειας από τους δικαιούχους. Στην παρουσίαση περιέχεται περιεχόμενο από τις ακόλουθες πηγές:

- Array mic, IMAGE Management LLC, Copyright © 1992-2009 CompanyLongName, 1501 Chester Road, Raleigh, NC 27608
- KURZWEIL Accelerating Intelligence, www.kurzweilai.net
- Ίδρυμα Μείζονος Ελληνισμού, Virtual Reality, www.fhw.gr

Σημείωμα Χρήσης Έργων Τρίτων (2 από 2)

Η δομή και οργάνωση της παρουσίασης, καθώς και το υπόλοιπο περιεχόμενο, αποτελούν πνευματική ιδιοκτησία του συγγραφέα και του Πανεπιστημίου Αθηνών και διατίθενται με άδεια Creative Commons Αναφορά Μη Εμπορική Χρήση Παρόμοια Διανομή Έκδοση 4.0 ή μεταγενέστερη.