

Αλληλεπίδραση Ανθρώπου-Υπολογιστή

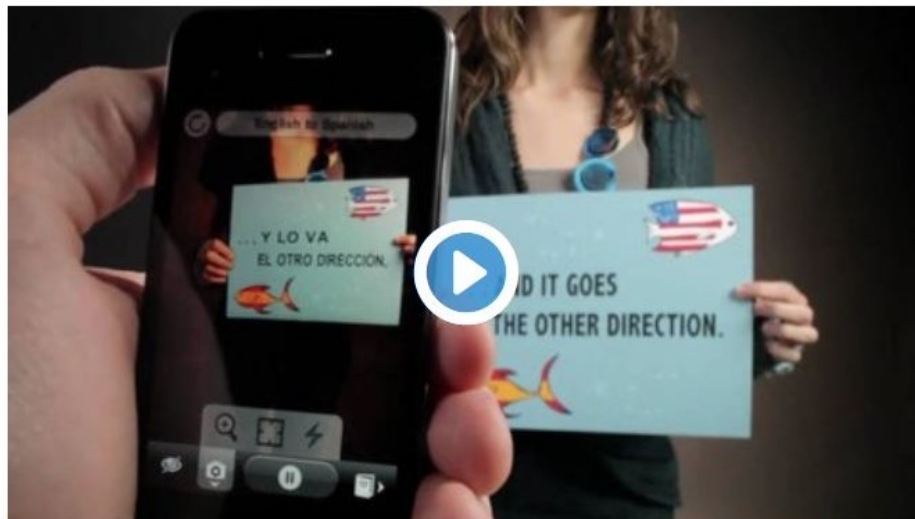
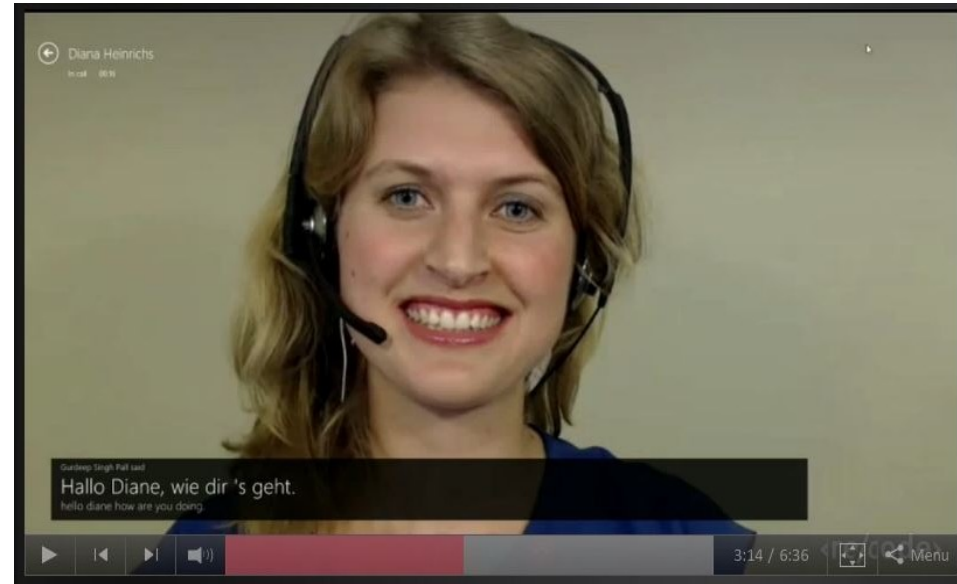
B1. Εισαγωγή στην τεχνολογία φυσικής γλώσσας

(2023–24)

Ίων Ανδρουτσόπουλος

<http://www.aueb.gr/users/ion/>

Μηχανική μετάφραση



<https://www.microsoft.com/en-us/research/video/speech-recognition-breakthrough-for-the-spoken-translated-word-short/>
<https://www.youtube.com/watch?v=RuAp92wW9bg>
<https://www.youtube.com/watch?v=h2OfQdYrHRs>

Εξόρυξη γνώμης

The screenshot shows a search interface for 'Apple Mac mini'. At the top, there is a search bar containing the text 'Apple Mac mini' and a blue 'GO' button. Below the search bar, there are four rows of sentiment analysis results, each represented by a white rounded rectangle with a vertical line separating the text from the star rating. The first row is for 'money, price, cost, ...' with a 5-star rating. The second row is for 'ram, memory, ...' with a 3-star rating. The third row is for 'design, color, feeling, ...' with a 4-star rating. The fourth row is for 'extras, keyboard, screen, ...' with a 2-star rating.

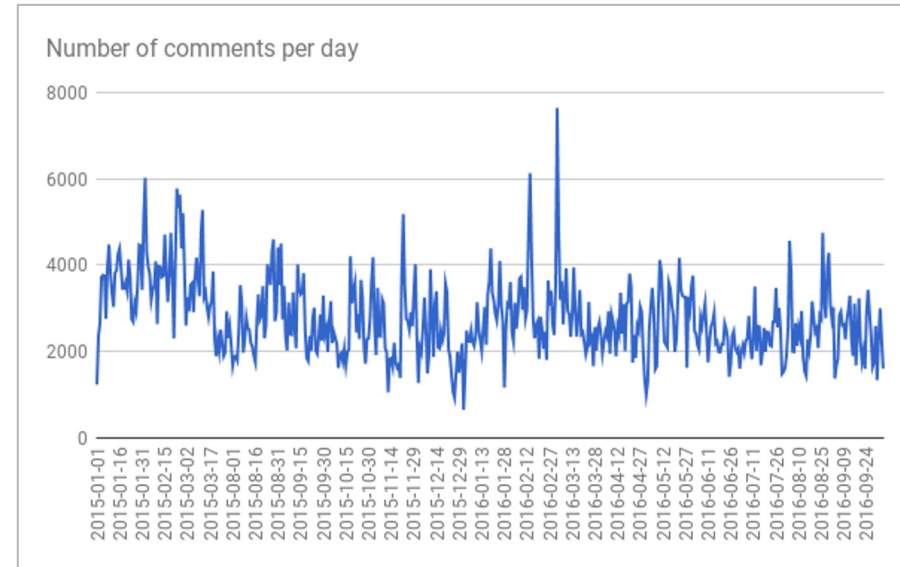
Aspect	Rating
money, price, cost, ...	5 stars
ram, memory, ...	3 stars
design, color, feeling, ...	4 stars
extras, keyboard, screen, ...	2 stars

- Από **κριτικές πελατών** (<http://alt.qcri.org/semEval2016/task5/>).
- Από **tweets** (<http://alt.qcri.org/semEval2016/task4/>), **FB updates**, **blogs**, **εφημερίδες**, **ραδιοφωνικές/τηλεοπτικές εκπομπές**, ...

Εντοπισμός «τοξικών» αναρτήσεων

A moderation panel assists the moderators to detect abusive comments, and leads to quicker publication of non-abusive comments.

Highlighting suspicious words using an RNN with self-attention.



Moderation Panel

Go	and	hang	yourself	!						85%		
You	are	ignorant	and	vandal	!	Stop	it	!		88%		
Hello	there	try	to	relax						0%		
Thanks	.	Please	go	f#\$@	yourself	.	Ty	!		85%		

Εξαγωγή πληροφοριών από έγγραφα

THIS AGREEMENT is made the 15th day of October 2009
(The “Effective Date”) BETWEEN:

- (1) **Sugar 13 Inc.**, a corporation whose office is at James House, 42-50 Bond Street, London, EW2H TL (“**Sugar**”);
- (2) **E2 UK Limited**, a limited company whose registered office is at 260 Bathurst Road, Yorkshire, SL3 4SA (“**Provider**”).

RECITALS:

- A. The Parties wish to enter into a framework agreement which will enable Sugar, from time to time, to [...]
- B. [...]

NO THEREFORE IT IS AGREED AS FOLLOWS:

ARTICLE I - DEFINITIONS

- “**Sugar**” shall mean: Sugar 13 Inc.
“**Provider**” shall mean: E2 UK Limited
“**1933 Act**” shall mean: **Securities Act of 1933**

ARTICLE II - TERMINATION

The Service Period will be for **five (5) years** from the Effective Date (The “**Initial Term**”). The agreement is considered to be terminated in **October 16, 2014**.

ARTICLE III - PAYMENT - FEES

During the service period monthly payments should occur. The estimated fees for the Initial Term are **£154,800**.

ARTICLE IV - GOVERNING LAW

This agreement shall be governed and construed in accordance with the **Laws of England & Wales**. Each party hereby irrevocably submits to the exclusive jurisdiction of the courts sitting in **Northern London**.

IN WITNESS WHEREOF, the parties have caused their respective duly authorized officers to execute this Agreement.

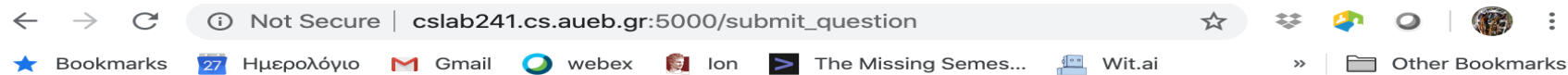
BY: George Fake
Authorized Officer
Sugar 13 Inc.

BY: Olivier Giroux
CEO
E2 UK LIMITED

Extract start/end dates,
durations, contractors,
amount, legal references,
jurisdiction etc.

- I. Chalkidis, I. Androutsopoulos and A. Michos, “Extracting Contract Elements”, ICAIL 2017, <http://nlp.cs.aueb.gr/pubs/icail2017.pdf>.
- I. Chalkidis and I. Androutsopoulos, “A Deep Learning Approach to Contract Element Extraction”, JURIX 2017, <http://nlp.cs.aueb.gr/pubs/jurix2017.pdf>.

Συστήματα ερωταποκρίσεων



Results for the question: How is Covid-19 transmitted?

Title: Coronavirus (COVID-19) Outbreak: What the Department of Radiology Should Know

Date: 2020-02-19 || Section: PRECAUTIONS FOR RADIOLOGY DEPARTMENT PERSONNEL
Available on: [PMID: 32092296](#)
Available on: [Doi : 10.1016/j.jacr.2020.02.008](#)

Coronavirus (COVID-19) Outbreak: What the Department of Radiology Should Know

Radiographers are among the first-line health care workers who might be exposed to 2019 novel COVID-19. Diagnostic imaging facilities should have guidelines in place to manage individuals with known or suspected COVID-19 infection.

The novel COVID-19 is highly contagious and is believed to transmit mostly through respiratory droplets, but there is uncertainty as to whether the virus can be transmitted by touching a surface or an item that is contaminated (ie, a fomite). A thorough understanding of the routes of virus transmission will be essential for patients' and health care professionals' safety.

Droplets have the greatest risk of transmission within 3 ft (91.44 cm), but they may travel up to 6 ft (183 cm) from their source [19] .

For the purpose of diagnostic imaging in individuals with NCIP, whenever possible, portable radiographic equipment should be used to limit transportation of patients.

On the basis of experience with SARS, the use of a satellite radiography center and dedicated radiographic equipment can decrease the risk for transmission from known infected individuals.

If a patient needs to be transported to the radiology department, he or she should wear a surgical mask during transport to and from the department.

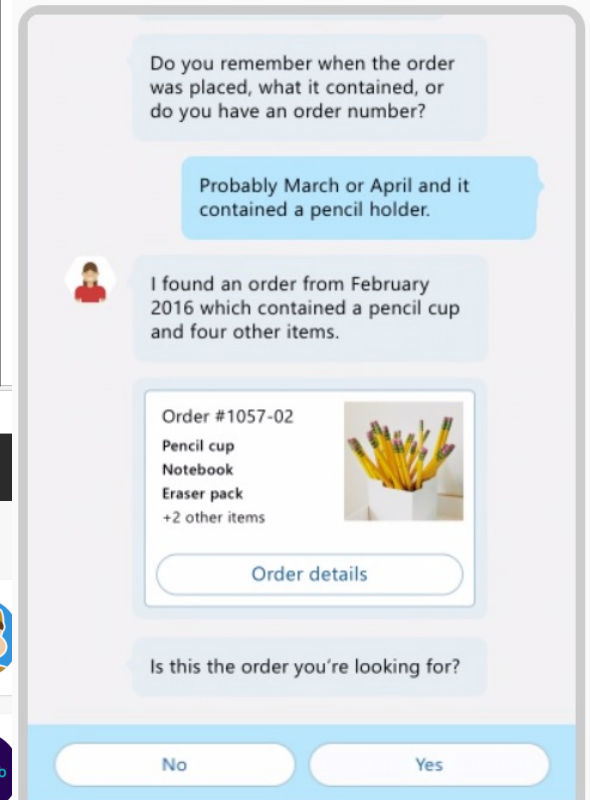
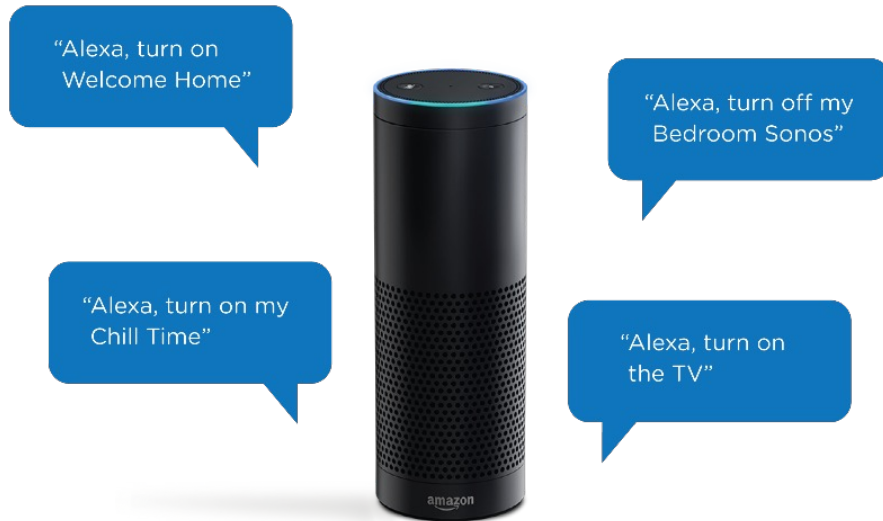
Title: Facemask shortage and the coronavirus disease (COVID-19) outbreak: Reflection on public health measures

Title: (Not Available)

Title: Epidemiological and clinical features of COVID-19 patients with and without pneumonia in Beijing, China

Title: Journal Pre-proof COVID-19 Spike-host cell receptor GRP78 binding site prediction COVID-19 Spike-host cell receptor GRP78 binding site prediction COVID-19 Spike -GRP78 binding site prediction









Συστήματα διαλόγων



Microsoft

Bot Framework PREVIEW My bots Documentation Bot Directory Blog

Featured bots

 <p>Skyscanner Skyscanner The world's travel search engine</p>	 <p>AzureBot Microsoft Be more productive with your own Microsoft Azure subscriptions using</p>	 <p>StubHub</p>
 <p>Jessie Humani PullString, Inc. Jessie Humani's life is in your hands. Will you help her make some</p>	 <p>SI x GameOn ... GameOn Technology Love the NFL? Add the SI x GameOn NFL bot to get unparalleled</p>	 <p>Solitaire Card Lion Solitaire is the most popular card game in the world.</p>
 <p>Baymax TrueCare24 Your personal healthcare companion. The service is available</p>	 <p>Cardea Powered by RingMD Meet Cardea, your personal medical aid. She will answer your health</p>	

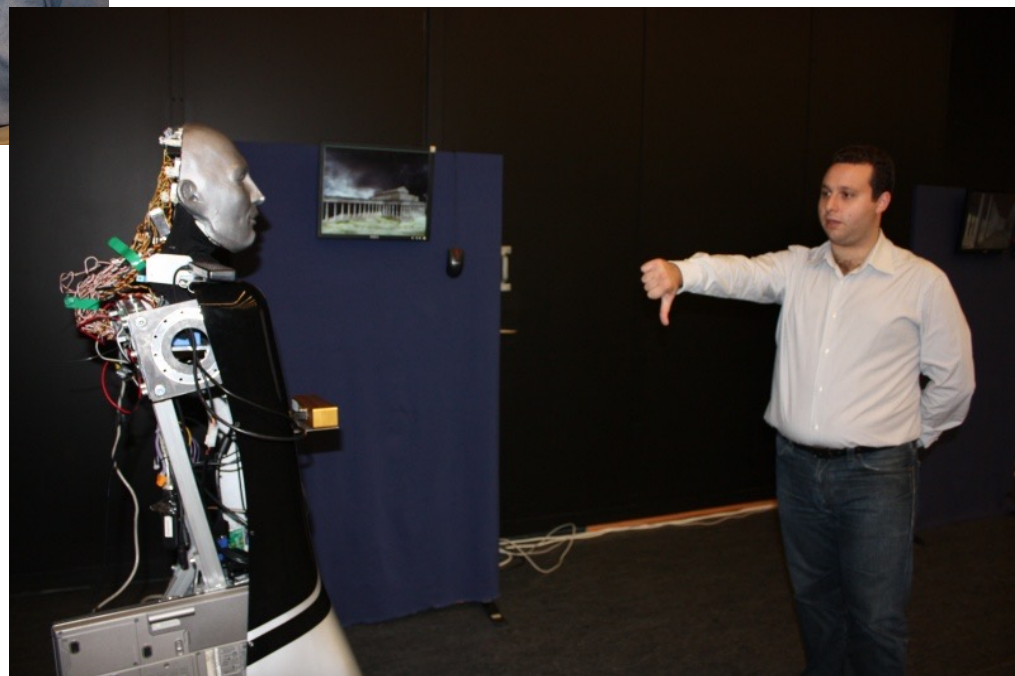
Διάλογοι με ρομποτικούς ξεναγούς



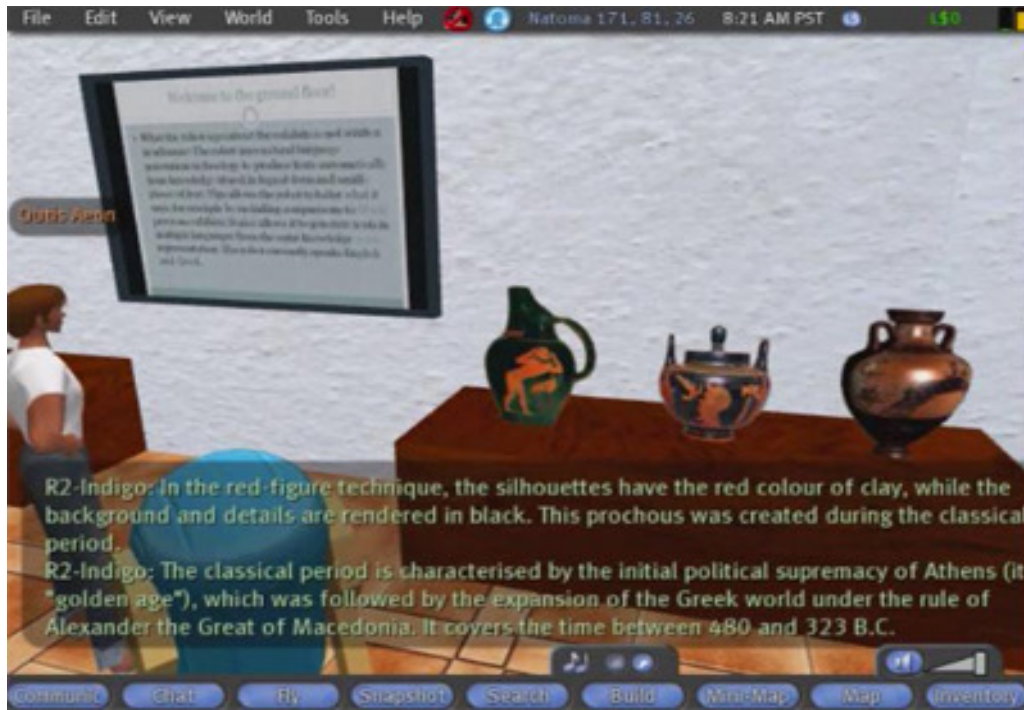
<https://www.ics.forth.gr/indigo/>

<https://www.youtube.com/watch?v=qCzBx4LzGak>

Η εικονοσκόπηση έγινε στον Ελληνικό Κόσμο του Ιδρύματος Μείζονος Ελληνισμού στη διάρκεια του IST INDIGO. Copyright εικόνων μνημείων ΙΜΕ (χρήση με την ευγενική παραχώρηση του ΙΜΕ). Τα κείμενα παράγονται από μια οντολογία OWL της Αγοράς της Αθήνας, που κατασκευάστηκε βάσει στοιχείων του ΙΜΕ.



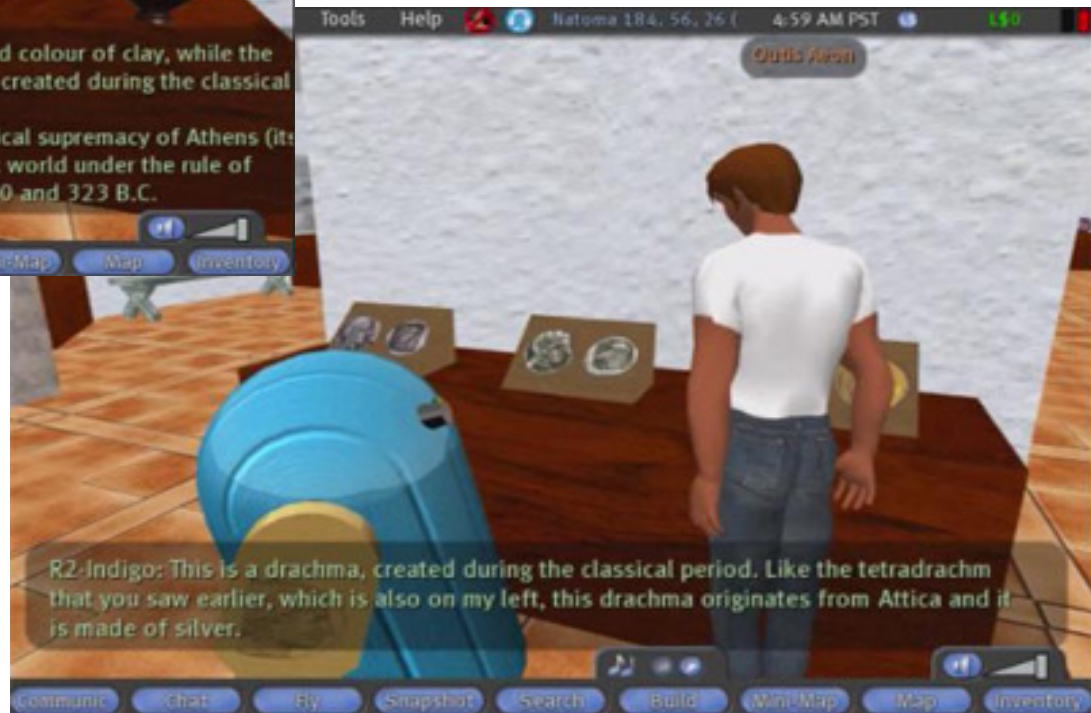
Αλληλεπίδραση σε εικονικούς κόσμους



Ξενάγηση σε εικονικό μουσείο στο Second Life (<http://secondlife.com/>).

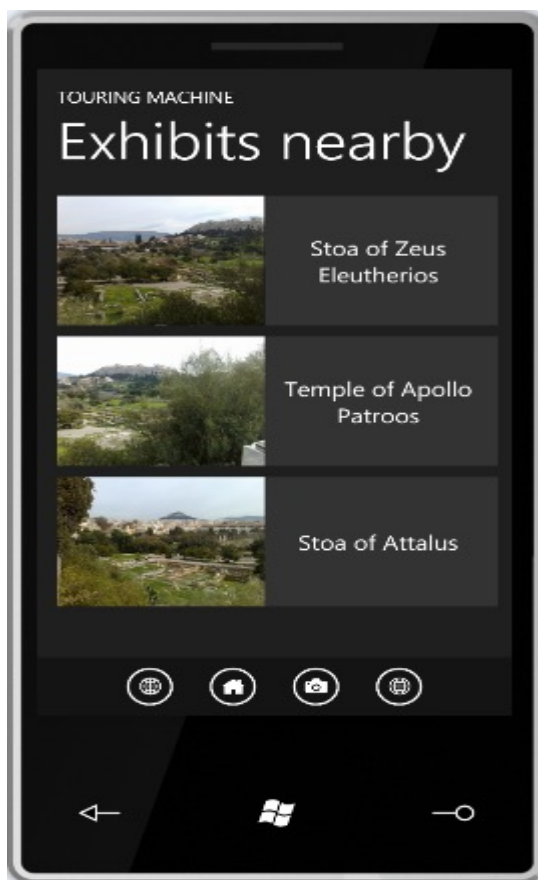
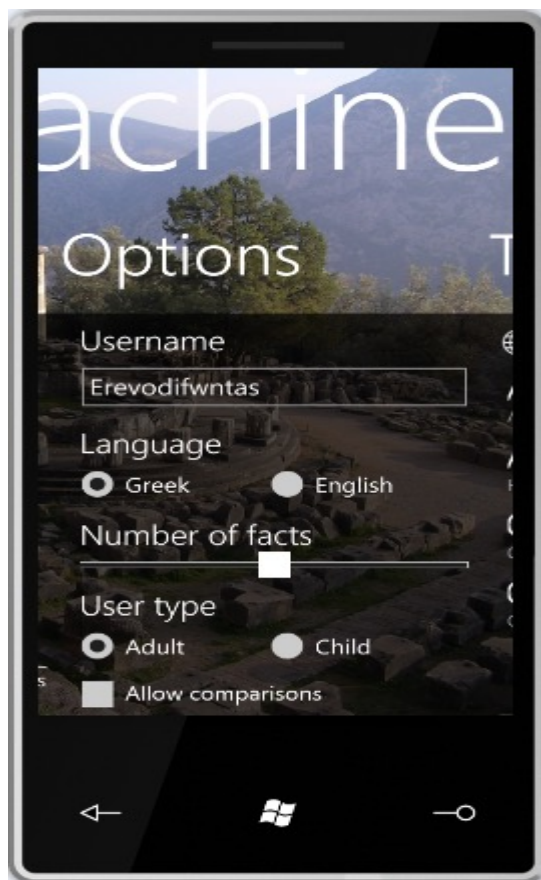
Βίντεο:

<https://vimeo.com/801099>



Copyright εικόνων εκθεμάτων ΙΜΕ (χρήση με την ευγενική παραχώρηση του ΙΜΕ). Τα κείμενα παράγονται από μια OWL μορφή της οντολογίας του ευρωπαϊκού έργου IST M-IRO (<https://cordis.europa.eu/project/rcn/55046/factsheet/en>). Η οντολογία βασίζεται σε στοιχεία του ΙΜΕ.

Παραγωγή περιγραφών αντικειμένων



Touring Machine (Γιώργος Καρακατσιώτης και Βαγγέλης Πτερνέας, Τμήμα Πληροφορικής, ΟΠΑ). Κέρδισε το ελληνικό Imagine Cup της Microsoft το 2010 και τον τελικό της κατηγορίας interoperability του διεθνούς Imagine Cup το 2011. Τα κείμενα παράγονται από μια οντολογία OWL της Αγοράς της Αθήνας, που κατασκευάστηκε στη διάρκεια του IST INDIGO (www.ics.forth.gr/indigo/) βάσει στοιχείων του IME.



I teach a 12 week long AI course. Can you suggest what I should cover per week?



Designing a 12-week AI course is a great endeavor. Below is a week-by-week breakdown of topics and concepts you could cover. Keep in mind that the course can be adjusted based on the prior knowledge of your students and the depth you want to go into each topic. Here's a general outline:



Week 1: Introduction to AI

- What is AI? History and milestones.
- Types of AI: Narrow (Weak) vs. General (Strong) AI.
- AI in popular culture and misconceptions.

Week 2: Machine Learning Fundamentals

- Introduction to Machine Learning (ML).
- Supervised, Unsupervised, and Reinforcement Learning.
- Data preprocessing and feature engineering.

Week 3: Supervised Learning

- Linear Regression.
- Logistic Regression.
- Evaluation metrics (MSE, MAE, accuracy, precision, recall, F1-score).

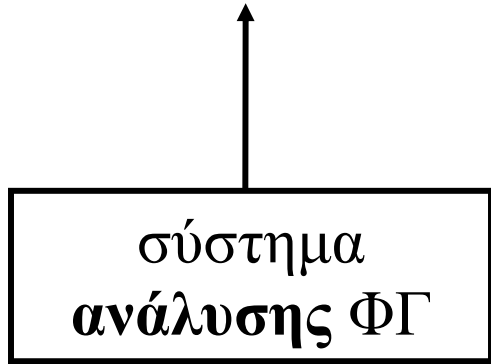
 Regenerate

Send a message



Ανάλυση και παραγωγή ΦΓ

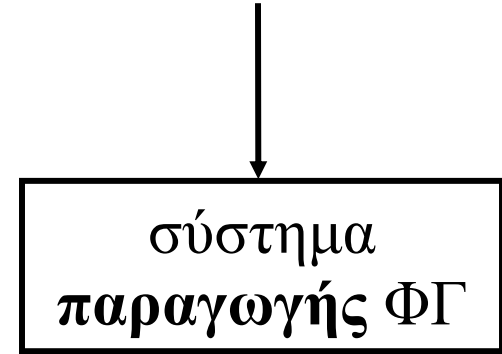
υπολογιστική παράσταση
(π.χ. λογική έκφραση, πίνακες
δεδομένων, στατιστικό μοντέλο)



κείμενο

(π.χ. ερώτηση, άρθρο
εφημερίδας, σώμα κειμένων)

υπολογιστική παράσταση
(π.χ. βάση δεδομένων, μετρήσεις
αισθητήρων, οντολογία)



κείμενο

(π.χ. περιγραφή αντικειμένου,
δελτίο καιρού, εγχειρίδιο)

Αλλά υπάρχουν και συστήματα που δημιουργούν **κείμενο από κείμενο** (text-to-text generation), π.χ. **μηχανική μετάφραση, παραγωγή περιλήψεων**.

Τεχνολογία φυσικής γλώσσας

- Αυτόματη **ανάλυση** και **παραγωγή γραπτών** ή **προφορικών εκφράσεων φυσικής γλώσσας**.
 - Αυτόματη **διόρθωση κειμένων**, έξυπνα **πληκτρολόγια**.
 - **Μηχανική μετάφραση**, πλέον και **προφορικού λόγου**.
 - **Διήθηση μηνυμάτων** (π.χ. φίλτρα spam, tweets).
 - **Εξαγωγή πληροφοριών** ή **εξόρυξη γνώμης** από κείμενα.
 - Συστήματα **ερωταποκρίσεων** (π.χ. αναζήτηση πληροφοριών σε άρθρα βιοϊατρικής).
 - Συστήματα **προφορικών διαλόγων** (π.χ. κρατήσεις εισιτηρίων, διάλογοι πλοήγησης σε αυτοκίνητο).
 - Συστήματα αυτόματης **παραγωγής κειμένων** (π.χ. περιγραφές προϊόντων ή εκθεμάτων σε πολλές γλώσσες).

Πολλοί συναφείς όροι

- **Επεξεργασία φυσικής γλώσσας** (natural language processing):
 - Κύριος στόχος η δημιουργία **υπολογιστικών συστημάτων** που να **κατανοούν** ή να **παράγουν γραπτές εκφράσεις** ΦΓ.
 - Συνήθως θεωρείται υποτομέας της Τεχνητής Νοημοσύνης.
- **Υπολογιστική γλωσσολογία** (computational linguistics):
 - Επίσης κυρίως **γραπτή φυσική γλώσσα**, κύριος στόχος η δημιουργία **υπολογιστικών μοντέλων γλωσσολογικών θεωριών**, αλλά στην πράξη **συνώνυμο της ΕΦΓ**.
 - Την τελευταία (τουλάχιστον) δεκαετία πιο κοντά στην Πληροφορική παρά στη Γλωσσολογία.
- **Γλωσσική τεχνολογία** (human language technology):
 - Λιγότερο καθιερωμένος όρος, συνήθως **περιλαμβάνει** και τεχνολογίες **αναγνώρισης** και **σύνθεσης φωνής**, έμφαση στη δημιουργία **χρήσιμων υπολογιστικών συστημάτων**.

Δύο είδη τεχνολογίας

Μέθοδοι που βασίζονται σε **χειρωνακτικά κατασκευασμένες παραστάσεις γνώσεων και κανόνες** (π.χ. λεξικά, κανόνες γραμματικών).

Δυσκολία κατασκευής και συντήρησης των παραστάσεων και κανόνων, **περιορισμένη κάλυψη** γλωσσικών φαινομένων.

Δεν απαιτούνται δεδομένα εκπαίδευσης αλλά ίσως **απαιτούνται γνώσεις γλωσσολογίας**.

Ευκολότερη ερμηνεία/κατανόηση της λειτουργίας και των **συμπερασμάτων** του συστήματος.

Μειούμενη χρήση τα τελευταία χρόνια αλλά **ακόμα σε χρήση**, π.χ. σε **διαχειριστές διαλόγων**.

Μέθοδοι που χρησιμοποιούν **μηχανική μάθηση**, τα τελευταία χρόνια **βαθιά μάθηση** (deep learning).

Το σύστημα **μαθαίνει** από προηγούμενα **παραδείγματα εκπαίδευσης και άλλα δεδομένα** (π.χ. σώματα κειμένων). Συνήθως **καλύτερες επιδόσεις και κάλυψη**.

Απαιτούνται μεγάλα σύνολα δεδομένων. Συχνά απαιτείται **δύσκολη/ακριβή χειρωνακτική επισημείωση** τουλάχιστον κάποιων **παραδειγμάτων εκπαίδευσης**.

Δυσκολότερη (ή αδύνατη) **ερμηνεία/κατανόηση** της λειτουργίας και των **συμπερασμάτων**.

Ραγδαία αυξανόμενη χρήση τα τελευταία χρόνια.

Βιβλιογραφία

- Όσοι ενδιαφέρονται ιδιαίτερα για την επεξεργασία φυσικής γλώσσας (και φωνής) αξίζει να μελετήσουν σταδιακά το εξαιρετικό βιβλίο «Speech and Language Processing» των D. Jurafsky and J.H. Martin, 2^η έκδοση, Prentice Hall, 2008.
 - Υπάρχει και στη βιβλιοθήκη του ΟΠΑ.
 - Διατίθεται ελεύθερα η υπό προετοιμασία 3^η έκδοση. Βλ. <http://web.stanford.edu/~jurafsky/slp3/>.
- Πολύ καλή εισαγωγή στη βαθιά μάθηση (deep learning) για επεξεργασία φυσικής γλώσσας είναι το βιβλίο του Y. Goldberg «Neural Network Models for Natural Language Processing», Morgan & Claypool Publishers, 2017.
- Δείτε προαιρετικά τις διαφάνειες του μεταπτυχιακού μαθήματος «Επεξεργασία Φυσικής Γλώσσας».
 - <https://eclass.aueb.gr/courses/INF210/>.

