



Πολιτισμική Πληροφορική

Ανάλυση
των τεκμηρίων

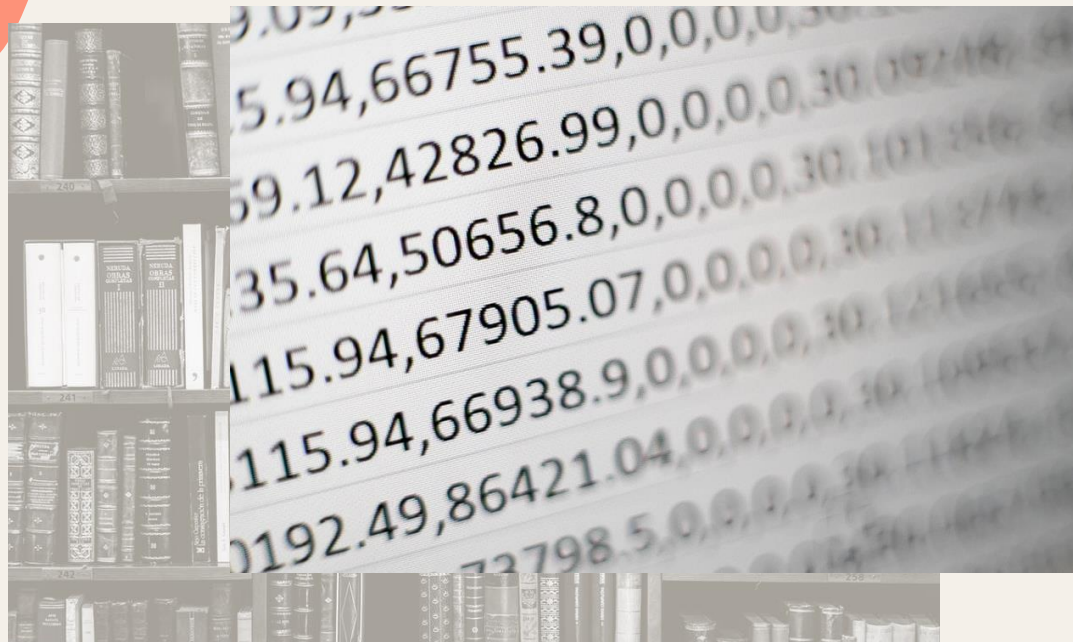
kkarpou@panteion.gr - @kkarpou

Κύκλος ζωής σε ένα έργο DH



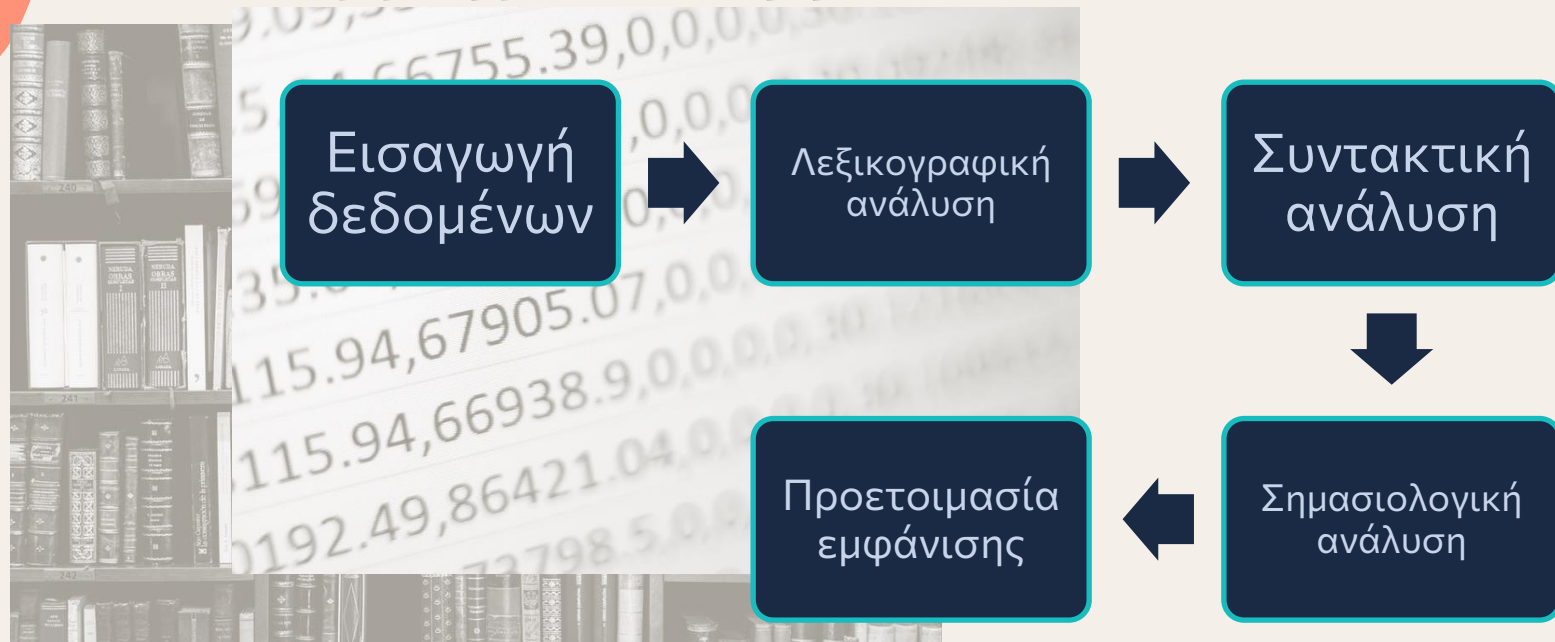
Πρωτογενή τεκμήρια

Κύκλος ζωής σε ένα έργο DH



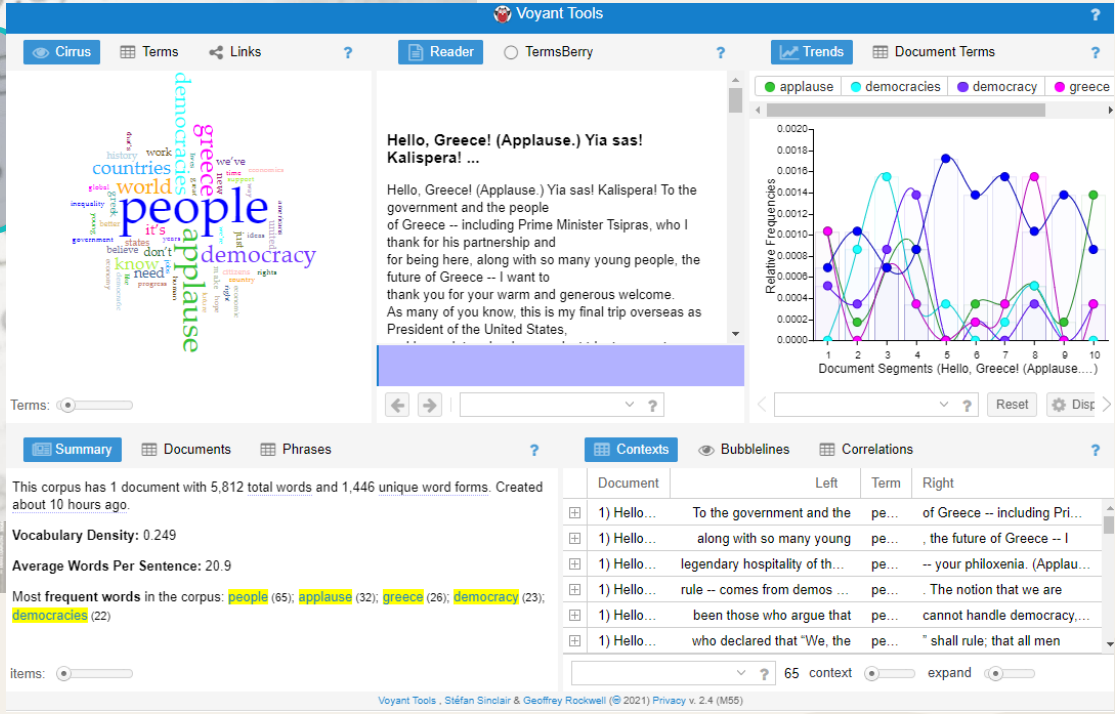
Δομημένα δεδομένα

Κύκλος ζωής σε ένα έργο DH

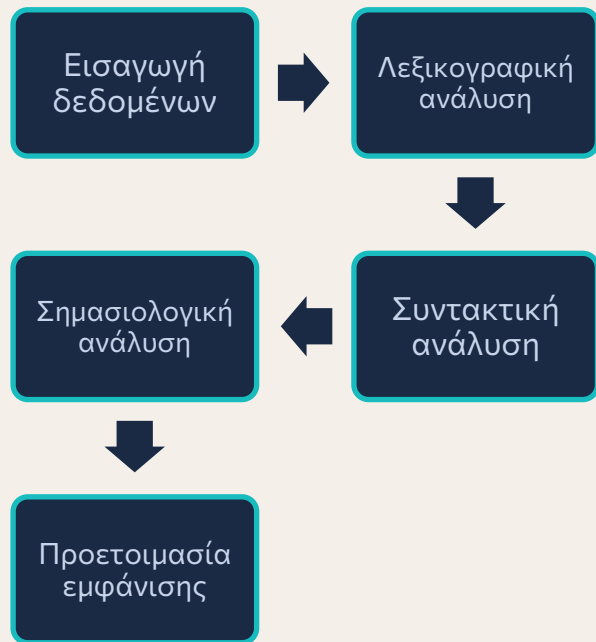


Υπηρεσίες επεξεργασίας

Εισαγωγή
δεδομένων

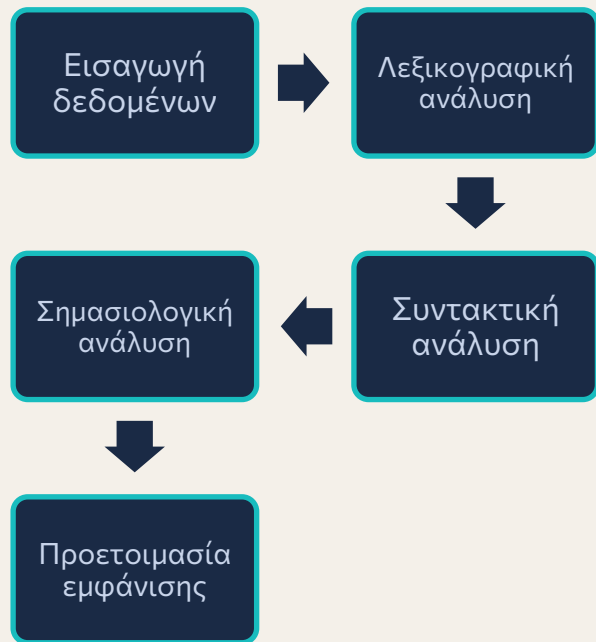


Υπηρεσίες επεξεργασίας



- Η συνολική διαδικασία μπορεί να απαρτίζεται από επιμέρους διεργασίες
 - Η καθεμιά ολοκληρώνεται από ξεχωριστό λογισμικό
 - Η επιλογή εξαρτάται από τη φύση των δεδομένων
 - Π.χ. εικόνες, κείμενο (εδώ), 3D αντικείμενα, κλπ.
 - Αλλά και τη δομή τους
- Αρτιότερη δομή → ευκολότερη επεξεργασία!

Υπηρεσίες επεξεργασίας



- Μπορεί να περιλαμβάνουν (ενδεικτικά)
 - Ανάλυση κειμένου (text analysis)
 - Ανάλυση εικόνων (image analysis)
 - Οπτικοποίηση πληροφοριών (visualization)
 - Ανάλυση δικτύων (network analysis)
 - Χωρική ανάλυση (GIS)
- Άλλο;
 - Ανάλογα με τα πιθανά ερωτήματα

Ανάλυση κειμένου

- Close reading
 - Jänicke, S., Franzini, G., Cheema, M. F., & Scheuermann, G. (2015, May). On Close and Distant Reading in Digital Humanities: A Survey and Future Challenges. In *EuroVis (STARs)* (pp. 83-103).
 - “the **careful, sustained interpretation** of a brief passage of a text. A close reading emphasizes the **single** and the **particular** over the **general**, effected by **close attention to individual words**, the **syntax**, the **order** in which the sentences unfold ideas, as well as formal structures. A truly attentive close reading of a two-hundred-word poem might be thousands of words long without exhausting the possibilities for observation and insight.”
(Wikipedia)

Ανάλυση κειμένου

- Close reading

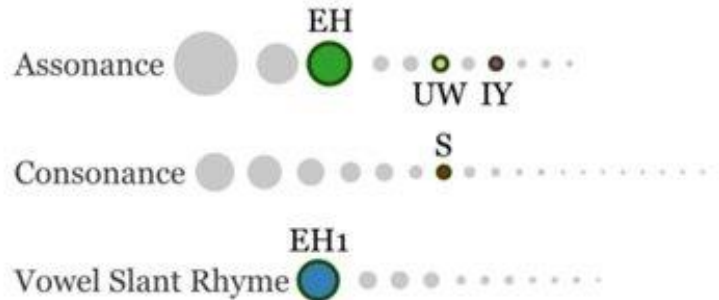
Night

Louise Bogan

The cold remote islands
And the blue estuaries
Where what breathes, breathes
The restless wind of the inlets,
And what drinks, drinks
The incoming tide;
Where shell and weed
Wait upon the salt wash of the sea
And the clear nights of stars
Swing their lights westward
To set behind the land;

Rhymes

Character Clusters



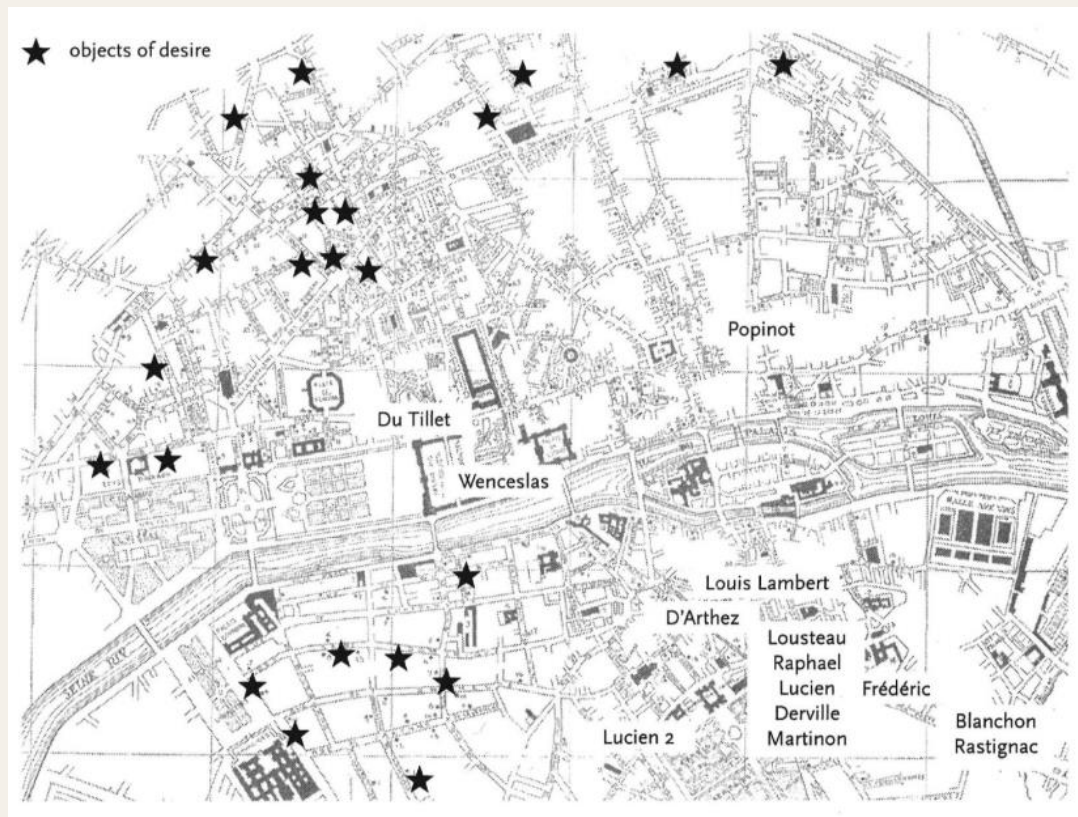
Ανάλυση κειμένου

- Close reading – μεθοδολογία
 - Επιλογή κειμένου ή εδαφίου (passage)
 - Ανάγνωση του κειμένου (πρώτη επαφή με την έκφραση)
 - Ανάλυση του κειμένου (εις βάθος γλωσσική επεξεργασία)
 - Μορφοποίηση ερωτημάτων
 - Πρωτόλεια μορφή απαντήσεων
 - Αποκλεισμός απαντήσεων – περιγραφική θέση
 - Επέκταση των πιθανών θέσεων
 - Ανάπτυξη περιγράμματος

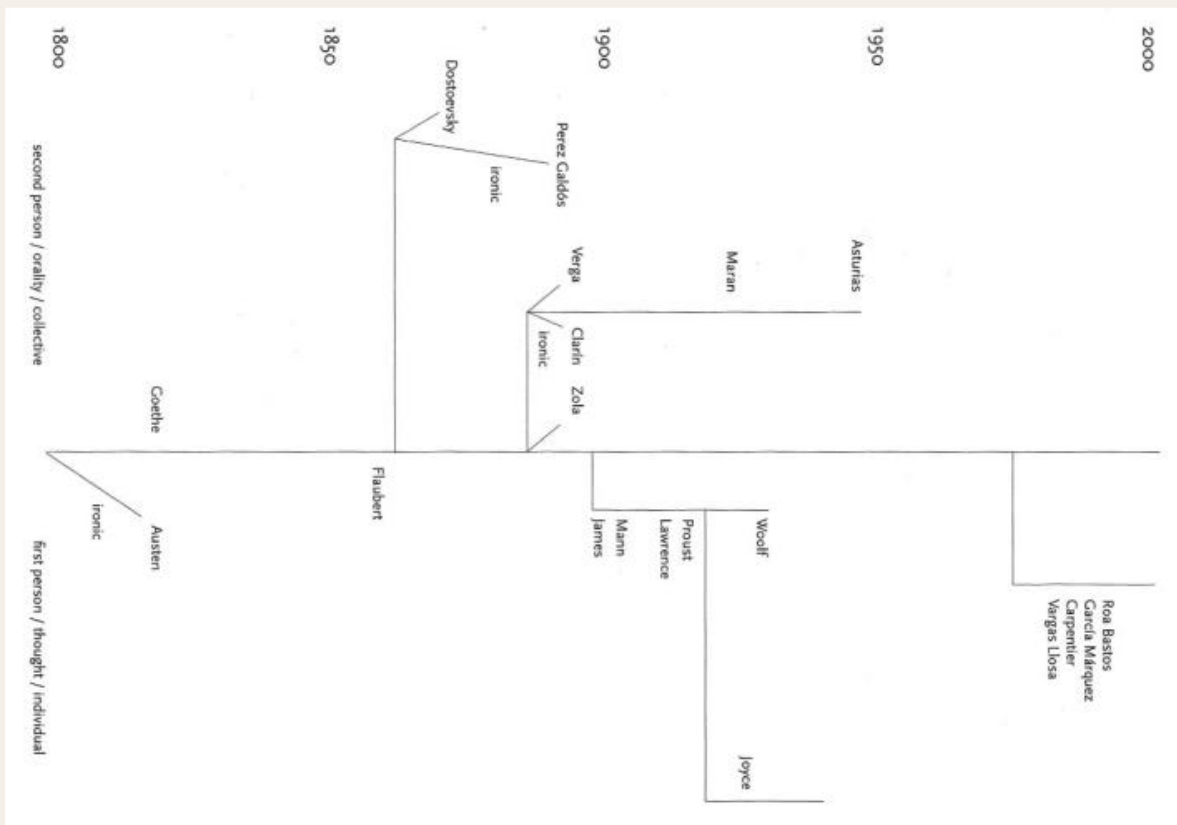
Ανάλυση κειμένου

- Distant reading
- Σε αντίθεση με το close reading που διατηρεί τη δομή του κειμένου → παρατήρηση καθολικών χαρακτηριστικών
- “we know how to read texts, now let’s learn how not to read them”, Moretti F. 2013. Distant reading. Verso, 2
- Συχνά συνοδεύεται από την οπτικοποίηση εννοιών, όπως
 - Γράφοι για την παρουσίαση των ειδών στα μυθιστορήματα
 - Χάρτες για τις γεωγραφικές έννοιες
 - Δενδρικές δομές για την ταξινόμηση αστυνομικών ιστοριών

Ήρωες σε Γαλλικά μυθιστορήματα



Ταξινόμηση ειδών σύγχρονης αφήγησης (1800-)



Μεθοδολογία distant reading

- Εντοπισμός προτύπων και εξέλιξης σε μεσαία και μεγάλη κλίμακα
 - Π.χ., παράγραφο
- Σύγκριση με όσα γνωρίζουμε για το είδος και την εποχή
- Χρήση ποσοτικών (αντικειμενικών) κριτηρίων

Επιλογή κειμένων

- Πρόκειται για κείμενο σε όλες τις γλώσσες;
- Περιλαμβάνει οπτικοποιημένο υλικό;
 - Φωτογραφίες, σκίτσα, χάρτες
- Έχει πολυμεσική/διαδραστική μορφή;
- Είναι αποτέλεσμα μεταγραφής;
- Για να υποστεί επεξεργασία, το κείμενο πρέπει να μπορεί να
 - Συλλεγεί σε ψηφιακή μορφή
 - Ταξινομηθεί
 - Αναλυθεί
 - Οπτικοποιηθεί

Επεξεργασία Φυσικής Γλώσσας

- Natural Language Processing (NLP)
- Κείμενα → μη δομημένα και δομημένα ταυτόχρονα!
 - Ως δεδομένα, μπορεί να μην έχουν δομή ή μεταδεδομένα
 - Ως γλώσσα, εμφανίζουν συντακτική δομή
- Lexical semantics: πώς κατανοούμε το νόημα των επιμέρους λέξεων
- Compositional semantics: πώς κατανοούμε το νόημα των προτάσεων

Επεξεργασία Φυσικής Γλώσσας

- Ουσιαστικά: αντικείμενα, τοπωνύμια, κύρια ονόματα
 - Βοηθούν στην ανάλυση του περιεχομένου
 - Named-entity recognition
- Ρήματα: δράσεις, ενέργειες
 - Περιγράφουν σχέσεις ανάμεσα σε οντότητες
- Επίθετα: περιγράφουν ουσιαστικά
- Επιρρήματα: χαρακτηρίζουν ρήματα

Παράδειγμα

- Έστω ότι το τεκμήριό μας είναι το παρακάτω κείμενο:
- “After the Ottoman conquest, the Parthenon was turned into a mosque in the early 1460s. On 26 September 1687, an Ottoman ammunition dump inside the building was ignited by Venetian bombardment during a siege of the Acropolis. The resulting explosion severely damaged the Parthenon and its sculptures. From 1800 to 1803, The 7th Earl of Elgin removed some of the surviving sculptures, now known as the Elgin Marbles, reportedly with the permission of the Turks of the Ottoman Empire.”
 - <https://en.wikipedia.org/wiki/Parthenon>
- Ποιες οντότητες μπορούμε να εντοπίσουμε αυτόματα;
- Ποιες άλλες πληροφορίες μάς δίνει το κείμενο;

Enter a sentence to extract named entities: it works well also on short texts.

Insert a or a of a newspaper/blog to analyze with Dandelion API:

After the Ottoman conquest, the Parthenon was turned into a mosque in the early 1460s. On 26 September 1687, an Ottoman ammunition dump inside the building was ignited by Venetian bombardment during a siege of the Acropolis. The resulting explosion severely damaged the Parthenon and its sculptures. From 1800 to 1803, The 7th Earl of Elgin removed some of the surviving sculptures, now known as the Elgin Marbles, reportedly with the permission of the Turks of the Ottoman Empire. ●

219

Language:

More Tags More Precision

ADVANCED+

Extract Entities

Try this examples:

Hint: too many entities?

If you have a long text and you need to select only the most important entities, take a look at the [new top_entities parameter](#).



Hey, did you:

- [get a free API key](#)
- [get started with the Entity Extraction API](#)
- [read API documentation](#)



Detected language: **English**

☐ OFF Show only top entities **NEW**

After the Ottoman conquest, the Parthenon was turned into a mosque in the early 1460s. On 26 September 1687, an Ottoman ammunition dump inside the building was ignited by Venetian bombardment during a siege of the Acropolis. The resulting explosion severely damaged the Parthenon and its sculptures. From 1800 to 1803, The 7th Earl of Elgin removed some of the surviving sculptures, now known as the Elgin Marbles, reportedly with the permission of the Turks of the Ottoman Empire.[15]

0 persons

1 work

0 organisations

3 places

2 events

7 concepts

Παράδειγμα

After the Ottoman conquest, the Parthenon was turned into a mosque in the early 1460s. On 26 September 1687, an Ottoman ammunition dump inside the building was ignited by Venetian bombardment during a siege of the Acropolis. The resulting explosion severely damaged the Parthenon and its sculptures. From 1800 to 1803, The 7th Earl of Elgin removed some of the surviving sculptures, now known as the Elgin Marbles, reportedly with the permission of the Turks of the Ottoman Empire.[15]

0 persons

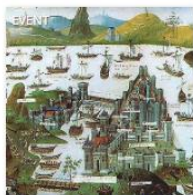
1 work

0 organisations

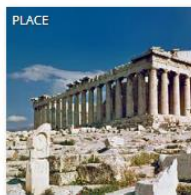
3 places

2 events

7 concepts



Fall of Constantinople



Parthenon



Mosque



Ammunition dump



Building



Republic of Venice



Bombardment



Morean War



Thomas Bruce, 7th Earl ...



Elgin Marbles

Μπορείτε να πειραματιστείτε με κείμενα στο
<https://dandelion.eu/semantic-text/entity-extraction-demo/>

Παράδειγμα

- “Beginning as early as AD 166–168, the Quadi and Marcomanni destroyed the main Roman town in the area, present-day Oderzo. This part of Roman Italy was again overrun in the early 5th century by the Visigoths and, some 50 years later, by the Huns led by Attila. The last and most enduring immigration into the north of the Italian peninsula, that of the Lombards in 568, left the Eastern Roman Empire only a small strip of coastline in the current Veneto, including Venice. The Roman/Byzantine territory was organized as the Exarchate of Ravenna, administered from that ancient port and overseen by a viceroy (the Exarch) appointed by the Emperor in Constantinople. Ravenna and Venice were connected only by sea routes, and with the Venetians' isolation came increasing autonomy. New ports were built, including those at Malamocco and Torcello in the Venetian lagoon. The tribuni maiores formed the earliest central standing governing committee of the islands in the lagoon, dating from c. 568.”

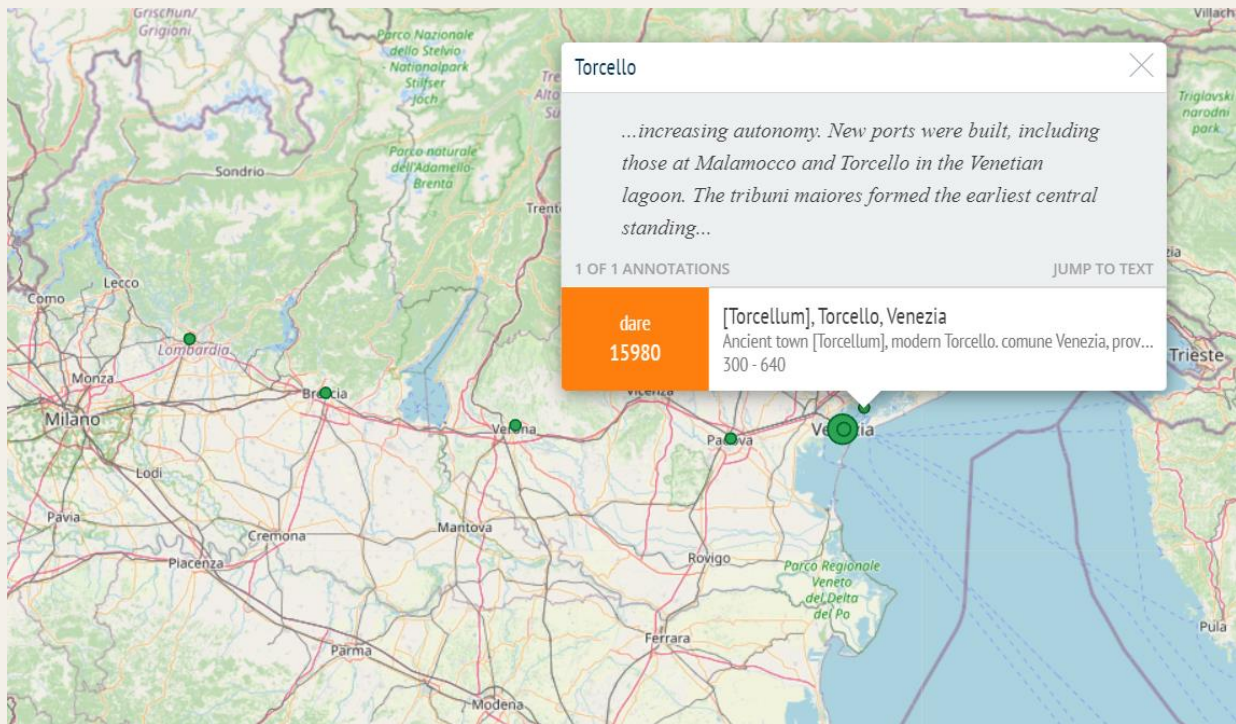
Recogito

- <https://recogito.pelagios.org/>

Beginning as early as AD 166–168, the **Quadi** and **Marcomanni** destroyed the main Roman town in the area, present-day **Oderzo**. This part of Roman Italy was again overrun in the early 5th century by the Visigoths and, some 50 years later, by the Huns led by **Attila**. The last and most enduring immigration into the north of the Italian peninsula, that of the **Lombards** in 568, left the Eastern Roman Empire only a small strip of coastline in the current Veneto, including Venice. The Roman/Byzantine territory was organized as the Exarchate of Ravenna, administered from that ancient port and overseen by a viceroy (the **Exarch**) appointed by the Emperor in Constantinople. **Ravenna** and **Venice** were connected only by sea routes, and with the Venetians' isolation came increasing autonomy. New ports were built, including those at Malamocco and **Torcello** in the Venetian lagoon. The tribuni maiores formed the earliest central standing governing committee of the islands in the lagoon, dating from c. 568.

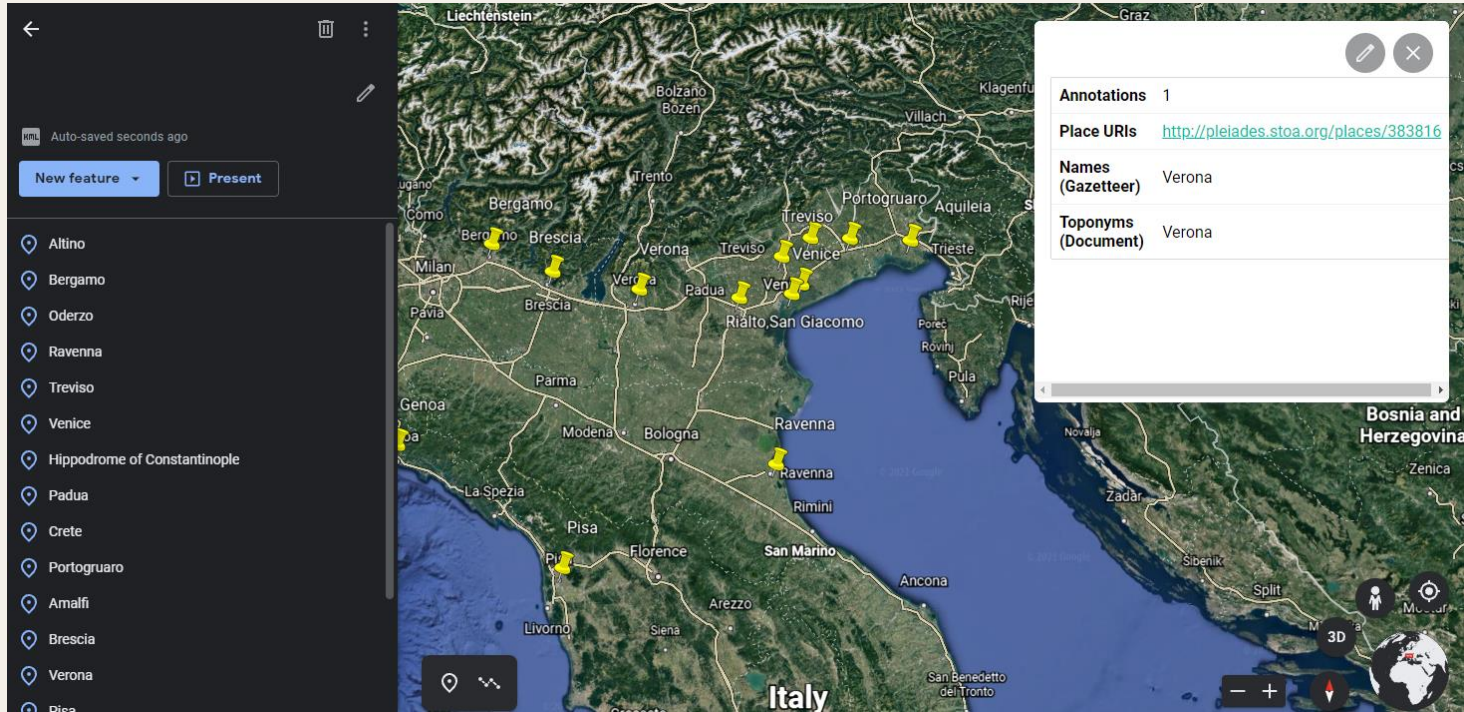
Recogito

- <https://recogito.pelagios.org/>



Recogito

- <https://recogito.pelagios.org/>



Μη δομημένο → ημιδομημένο

<p>VENICE</p>

<p>

Although no surviving historical records deal directly with the founding of

<placeName ref="http://sws.geonames.org/3164603" xml:id="recogito-bde7b8de-00da-4a9d-9cf8-15e13ba42d96" cert="high">Venice</placeName>

, tradition and the available evidence have led several historians to agree that the original population of

<placeName ref="http://sws.geonames.org/3164603" xml:id="recogito-f0e60979-b0d4-4036-9386-eea06439eb50" cert="high">Venice</placeName>

consisted of refugees—from nearby Roman cities such as

<placeName ref="http://pleiades.stoa.org/places/393473" xml:id="recogito-caf0e318-9d09-4fee-bcb2-1c4a2c33c28b" cert="high">Padua</placeName>

, <placeName ref="http://pleiades.stoa.org/places/187290" xml:id="recogito-8da90f9c-6dfd-45d9-919f-d22ff3157481" cert="high">Aquileia</placeName>

, <placeName ref="http://pleiades.stoa.org/places/393503" xml:id="recogito-79356f25-3cec-467e-bc07-a82c750fc4ee" cert="high">Treviso</placeName>

, <placeName ref="http://sws.geonames.org/3183163" xml:id="recogito-ed55b5b4-1a80-456f-81ff-c48d96b5dfec" cert="high">Altino</placeName>

, and <placeName ref="http://pleiades.stoa.org/places/240912" xml:id="recogito-7bb358e2-33bd-4569-a02e-09f61ec68804" cert="high">Concordia</placeName>

(modern

<placeName ref="http://sws.geonames.org/3170086" xml:id="recogito-107bda34-5f7d-471e-9d6d-c33c48b4478e" cert="high">Portogruaro</placeName>

), as well as from the undefended countryside—who were fleeing successive

<rs xml:id="recogito-c58da6f8-8d9d-411f-99d9-3c50ed2784df" type="event">waves of Germanic and Hun invasions</rs>

. This is further supported by the documentation on the so-called "apostolic families", the twelve founding families of

<placeName ref="http://sws.geonames.org/3164603" xml:id="recogito-34707b40-dc9a-4578-b331-81cb02448df5" cert="high">Venice</placeName>

who elected the

<persName xml:id="recogito-8d9f8615-0e38-4eb7-b9d7-fd6dc3b0db4a">first doge</persName>

, who in most cases trace their lineage back to Roman families. Some late Roman sources reveal the existence of fishermen, on the islands in the

original marshy lagoons, who were referred to as incolae lacunae ("lagoon dwellers"). The traditional founding is identified with the dedication of

the first church, that of

<placeName ref="http://dare.ht.lu.se/places/39589" xml:id="recogito-8753bdd6-e2ee-47ef-a9a8-1f0967d008c2" cert="high">San Giacomo</placeName>

on the islet of Rialto(Rivoalto, "High Shore")—said to have taken place at the stroke of noon on

<rs xml:id="recogito-2d9cee9b-03eb-4e2f-9e04-0cef56b2b389" type="event">25 March 421</rs>

(the Feast of the Annunciation).

</p>

Εντοπισμός χαρακτηριστικών σε ένα κείμενο

- <https://www.uclassify.com/>
- “In March 2011, archaeologists announced that they had discovered five metopes of the Parthenon in the south wall of the Acropolis, which had been extended when the Acropolis was used as a fortress. According to Eleftherotypia daily, the archaeologists claimed the metopes had been placed there in the 18th century when the Acropolis wall was being repaired. The experts discovered the metopes while processing 2,250 photos with modern photographic methods, as the white Pentelic marble they are made of differed from the other stone of the wall. It was previously presumed that the missing metopes were destroyed during the Morosini explosion of the Parthenon in 1687”, Wikipedia

Εντοπισμός χαρακτηριστικών σε ένα κείμενο

- Ηλικία συγγραφέα

65-100

43%

51-65

21%

18-25

12%

13-17

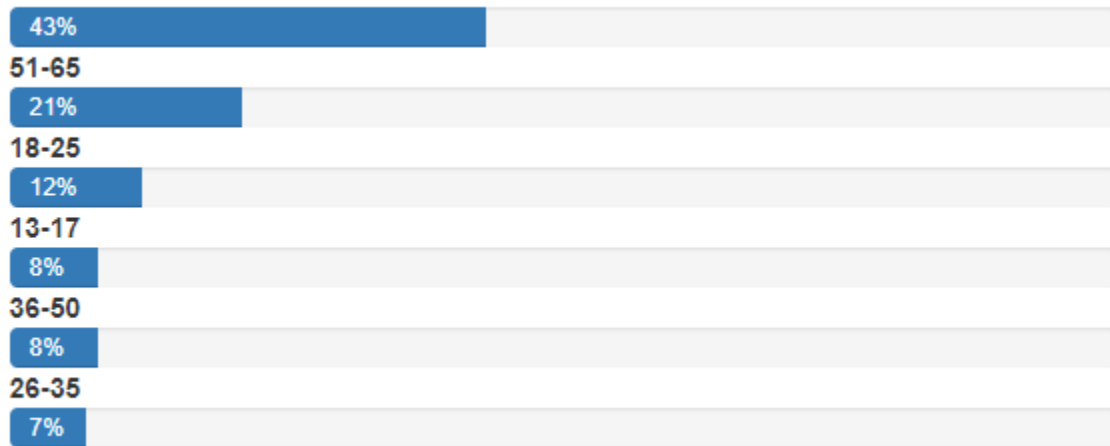
8%

36-50

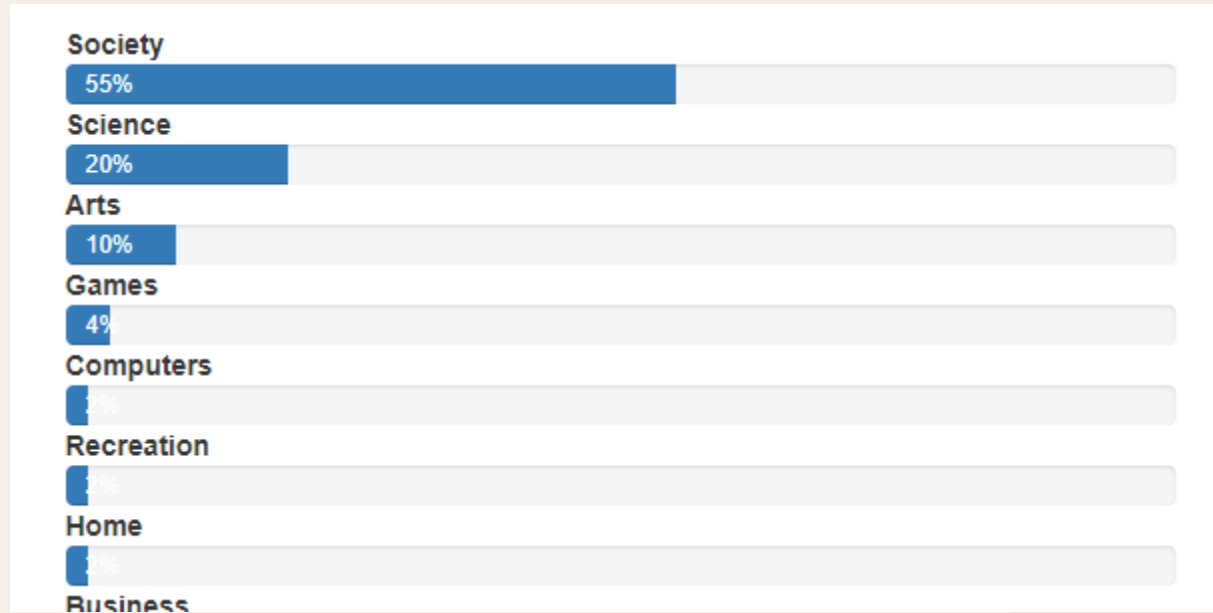
8%

26-35

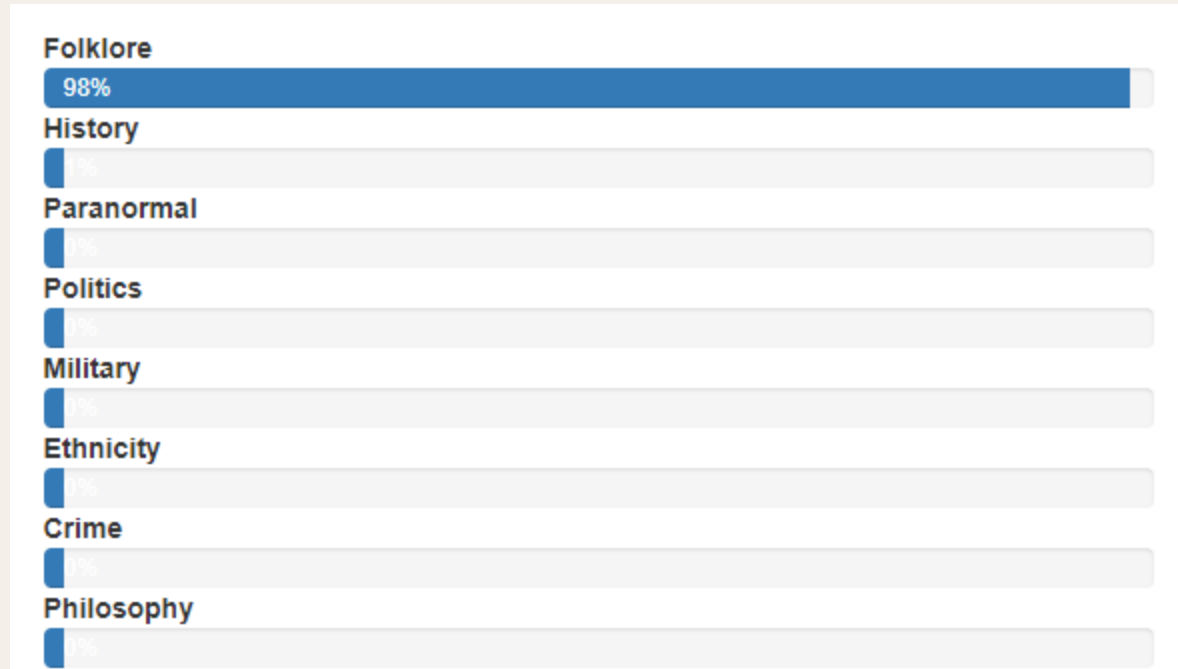
7%



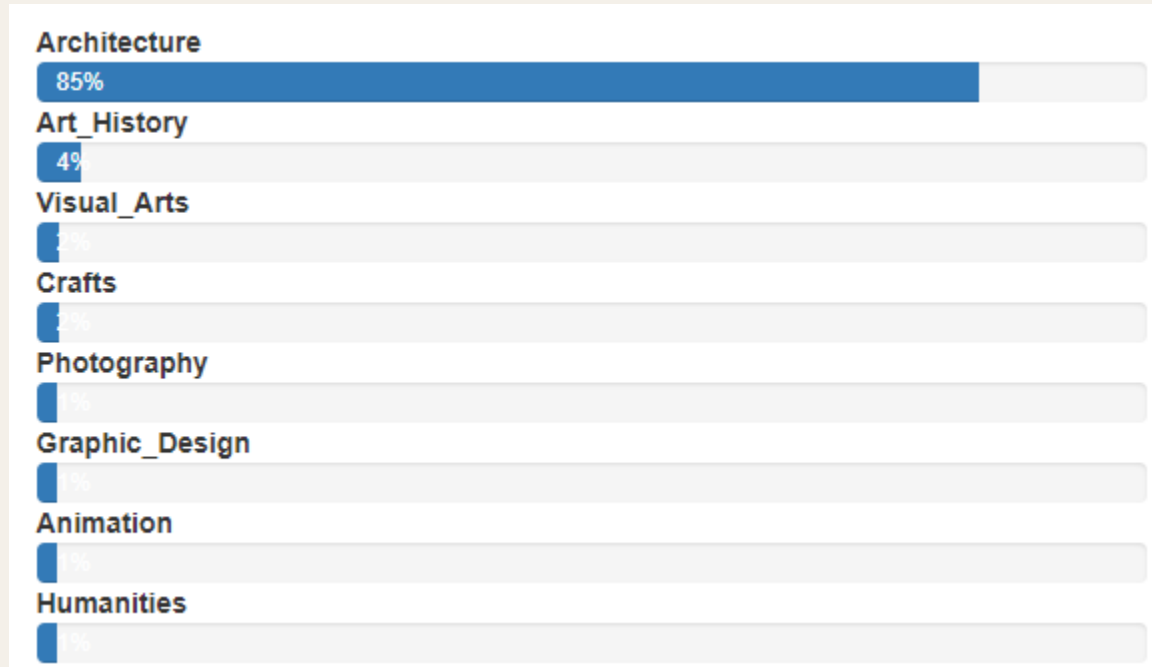
- Θεματολογία



- Θεματολογία σε σχέση με την κοινωνία



- Θεματολογία σε σχέση με τις τέχνες



Περιορισμοί στην ανάλυση κειμένων

- Συνήθως εργαζόμαστε με **δείγματα κειμένων** από μια συνολική συλλογή
 - Συχνά η πλήρης συλλογή δεν είναι καν διαθέσιμη (π.χ., αρχαίο δράμα)
- Η γλώσσα των κειμένων **δεν είναι πάντα δομημένη** όσο θα θέλαμε
 - Ποίηση, λογοτεχνία, τραγούδι
- Μπορεί το κείμενο να είναι γραμμένο σε κάποια **διάλεκτο**
- **Ηθικά** ζητήματα
 - Πνευματική ιδιοκτησία
 - Τι (πραγματικά) θέλει να πει ο ποιητής;
- Οι **μέθοδοι μηχανικής μάθησης** δεν είναι πάντα **αξιόπιστες**
 - Το αποτέλεσμα εξαρτάται από τα κείμενα με τα οποία εκπαιδεύτηκαν
 - Και από τα δεδομένα εκπαίδευσης που εισαγάγαμε (πιθανό bias)

Οπτική ανάλυση

Unusually, the artist, Taylor Swift, is not connecting with the audience through a direct mode of address. Her eyes, rather than addressing the audience focus on the album title 'Red'. Seeing as this is Swift's third studio album she no longer needs to have her entire face in view on a cover as her fans already know and support her. Well established artists often choose not to appear in album covers, and opt instead for cover art.

The shadow cast by her hat covers all of Taylor's face with the exception of her lips. Red lipstick has recently become a popular fashion choice and by supporting this look Taylor makes herself appeal to the mainstream public. This also helps emphasize her star image.

The same bold typography is used for the album title and the artists name, a common convention of album covers. Capitalization has been used to have a bold impact and the red colouring on the title 'RED' helps amplify the title and also creates a colour scheme, complimenting the shade of Taylor's lipstick.



The digipak follows a neutral colour scheme with the vibrant pop of red. A shallow depth of field has been used to separate the artist from the background and helps imply the significance of her. A basic cream colour has been used for the background, again to not distract from the artist. The colour red has been used throughout the digipak to visually echo the album title "RED".



Οπτική ανάλυση

- Μπορεί να γίνει χειροκίνητα από κάποιον ειδικό
- Πλέον και αυτόματα, με μεγάλη επιτυχία
 - Automatic image tagging
 - “the process by which a computer system automatically assigns metadata in the form of captioning or keywords to a digital image”, Wikipedia
- Budzise-Weaver, T. 2016. “Developing a Qualitative Coding Analysis of Visual Artwork for Humanities Research”. *Digital Humanities Quarterly*, vol. 10(4)
 - <http://www.digitalhumanities.org/dhq/vol/10/4/000275/000275.html>
- Manovich, L. 2015. “Data Science and Digital Art History”. *International Journal for Digital Art History*, no. 1 (June).
<https://doi.org/10.11588/dah.2015.1.21631>.

Τι μπορώ να εντοπίσω με την οπτική ανάλυση

- Συγκεκριμένους **τύπους αντικειμένων** σε μια εικόνα (π.χ., αυτοκίνητα ή ζώα)
- Συγκεκριμένους (ή όχι) **ανθρώπους**
- Εικόνες που έχουν **κοινά χαρακτηριστικά** μέσα σε ένα μεγάλο σώμα δεδομένων
 - Πίνακες σε μια πινακοθήκη, εξώφυλλα περιοδικών
- Ακόμα
 - Να **δημιουργήσω καινούριες εικόνες** που μοιάζουν με μια δεδομένη
 - Να **μεταφέρω το στυλ** μιας εικόνας σε μια άλλη (style transfer)

Τι μπορώ να εντοπίσω με την οπτική ανάλυση



Generated tags

English

Concepts

tower	72.75%
architecture	62.68%
tourism	50.49%
monument	46.85%
famous	46.68%
landmark	45.29%
travel	43.09%
city	37.53%
building	35.87%
tourist	35.48%

[↪ show me more tags](#)

Τι μπορώ να εντοπίσω με την οπτική ανάλυση



Generated tags

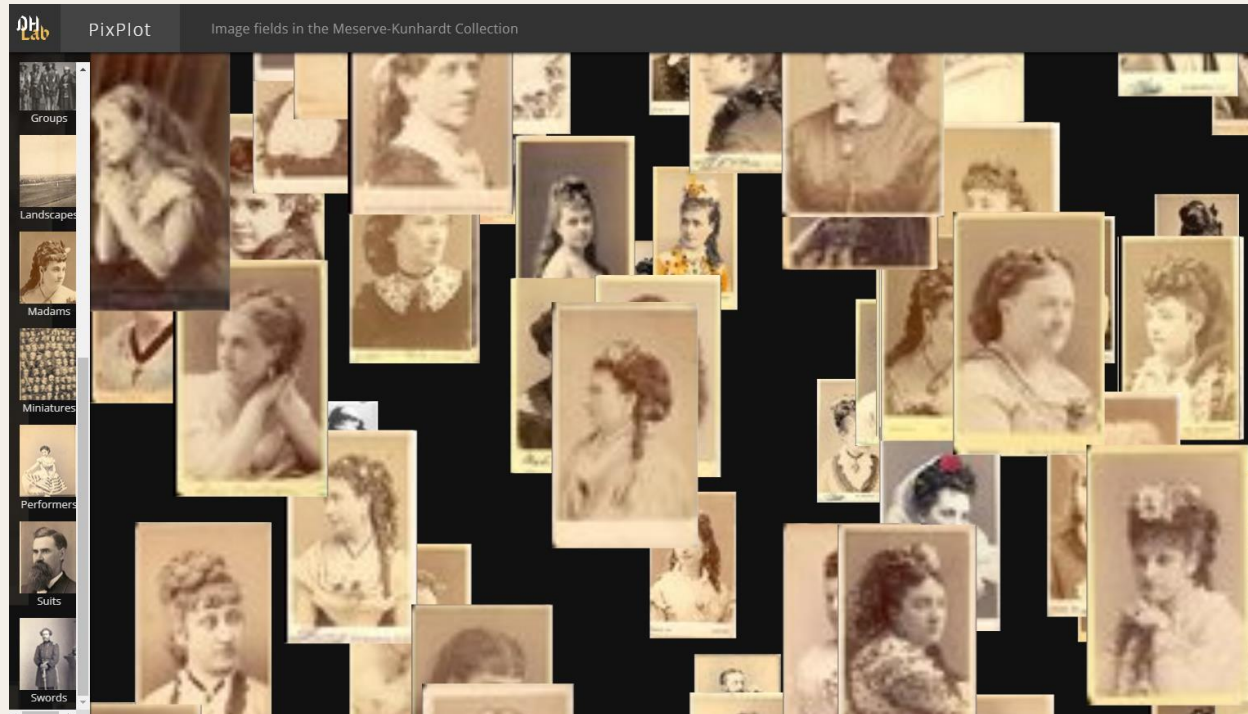
English

Concepts

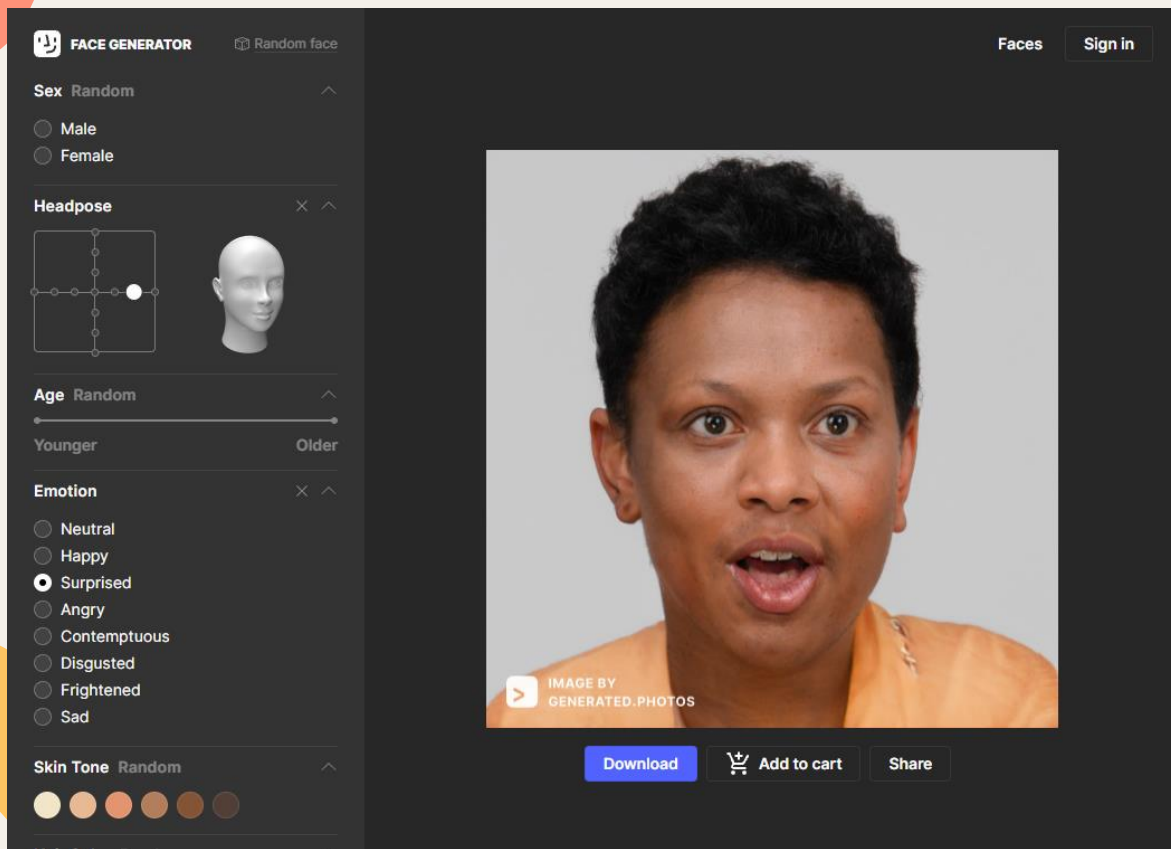
tower	72.75%
architecture	62.68%
tourism	50.49%
monument	46.85%
famous	46.68%
landmark	45.29%
travel	43.09%
city	37.53%
building	35.87%
tourist	35.48%
sky	34.64%
capital	30.45%
attraction	26.84%
symbol	26.35%
metal	24.24%
structure	22.84%
construction	22.32%
urban	21.93%

Εντοπισμός παρόμοιων εικόνων

- PixPlot, <https://dhlabs.yale.edu/projects/pixplot/>



Δημιουργία συνθετικών εικόνων



FACE GENERATOR

Sex Random

Male

Female

Headpose

Age Random

Younger

Emotion

Neutral

Happy

Surprised

Angry

Contemptuous

Disgusted

Frightened

Sad

Skin Tone Random

White

Light

Medium

Dark

Ethnic Traits Random

Epicanthus

Happy

Surprised

Angry

Contemptuous

Disgusted

Frightened

Sad

Skin Tone Random

White

Light

Medium

Dark

Hair Color Grey

White

Light

Medium

Dark

Black

Hair Length Random

Short

Medium

Long

Hair loss

Glasses Random

No

Reading glasses

Sunglasses

Makeup Random

Eyes

Lips

Ethnic Traits Random

Epicanthus

Download

Add to cart

Share

Faces

Sign in

IMAGE BY
GENERATED.PHOTOS

Δημιουργία συνθετικών εικόνων

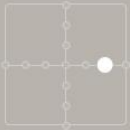
FACE GENERATOR

Sex Random

☐ Male

☐ Female

Headpose



Age Random

Emotion

☐ Neutral

☐ Happy

☒ Surprised

☐ Angry

☐ Contemptuous

☐ Disgusted

☐ Frightened

☐ Sad

Skin Tone Random

☐ Happy

☐ Surprised

☐ Angry

☐ Contemptuous

☒ Disgusted

☐ Frightened

☐ Sad

Skin Tone Random

☐ No

☐ Reading glasses

☐ Sunglasses

Makeup Random

☐ Eyes

☐ Lips

Ethnic Traits Random

☐ Epicanthus

Emotion

☐ Neutral

☒ Happy

☐ Surprised

☐ Angry

☐ Contemptuous

☐ Disgusted

☐ Frightened

☐ Sad

Skin Tone Fair

☐ No

☐ Reading glasses

☐ Sunglasses

Makeup Random

☐ Eyes

☐ Lips

Ethnic Traits

Faces

Sign in



IMAGE BY
GENERATED.PHOTOS

Download

Add to cart

Share

Style transfer

Fast Style Transfer Example

INPUT



[Upload an image](#)

[Use my webcam](#)

[Transfer My Image / Video](#)

Click this button to start transferring your own image or video.

STYLE



OUTPUT



Σύνδεσμοι

- F. Moretti, Graphs, Maps, Trees: Abstract Models for Literary History, Verso books, <https://www.versobooks.com/books/261-graphs-maps-trees>
- Voyant: see through your text, <https://voyant-tools.org/>
- Entity extraction, <https://dandelion.eu/semantic-text/entity-extraction-demo/>
- Recogito, <https://recogito.pelagios.org/>
- PixPlot, <https://dhlabs.yale.edu/projects/pixplot/>
- uClassify, <https://www.uclassify.com/>
- AI generated photos, <https://generated.photos/faces>
- Style transfer, https://yining1023.github.io/fast_style_transfer_in_ML5/#

Ερωτήσεις αυτοαξιολόγησης

- Έστω ότι εργάζεστε σε ένα έργο DH όπου τα τεκμήρια είναι κείμενα σχετικά με την επανάσταση του 1821. Περιγράψτε μια εργασία για καθένα από τα βήματα του κύκλου ζωής του έργου
- Αναφέρετε ένα ερώτημα που μπορούμε να απαντήσουμε με close reading και ένα με distant reading και δώστε μια εκτίμηση για τις απαντήσεις
- Από την εμπειρία σας, ποια επίπτωση μπορεί να έχει η χρήση ενός κειμένου σε περισσότερες από μια γλώσσες;
- Περιγράψτε ένα έργο DH στο οποίο θα είχε νόημα η εξαγωγή ενός χάρτη από τα κειμενικά τεκμήρια
- Περιγράψτε ένα έργο DH στο οποίο θα είχε νόημα η τεχνολογία style transfer



Πολιτισμική Πληροφορική

Ανάλυση
των τεκμηρίων

kkarpou@panteion.gr - @kkarpou