



# Σημασιολογικός και Κοινωνικός Ιστός

## Διάλεξη 05: Αναγνωριστικά

Γεώργιος Δημητρακόπουλος  
dimitrakopoulos@ionio.gr

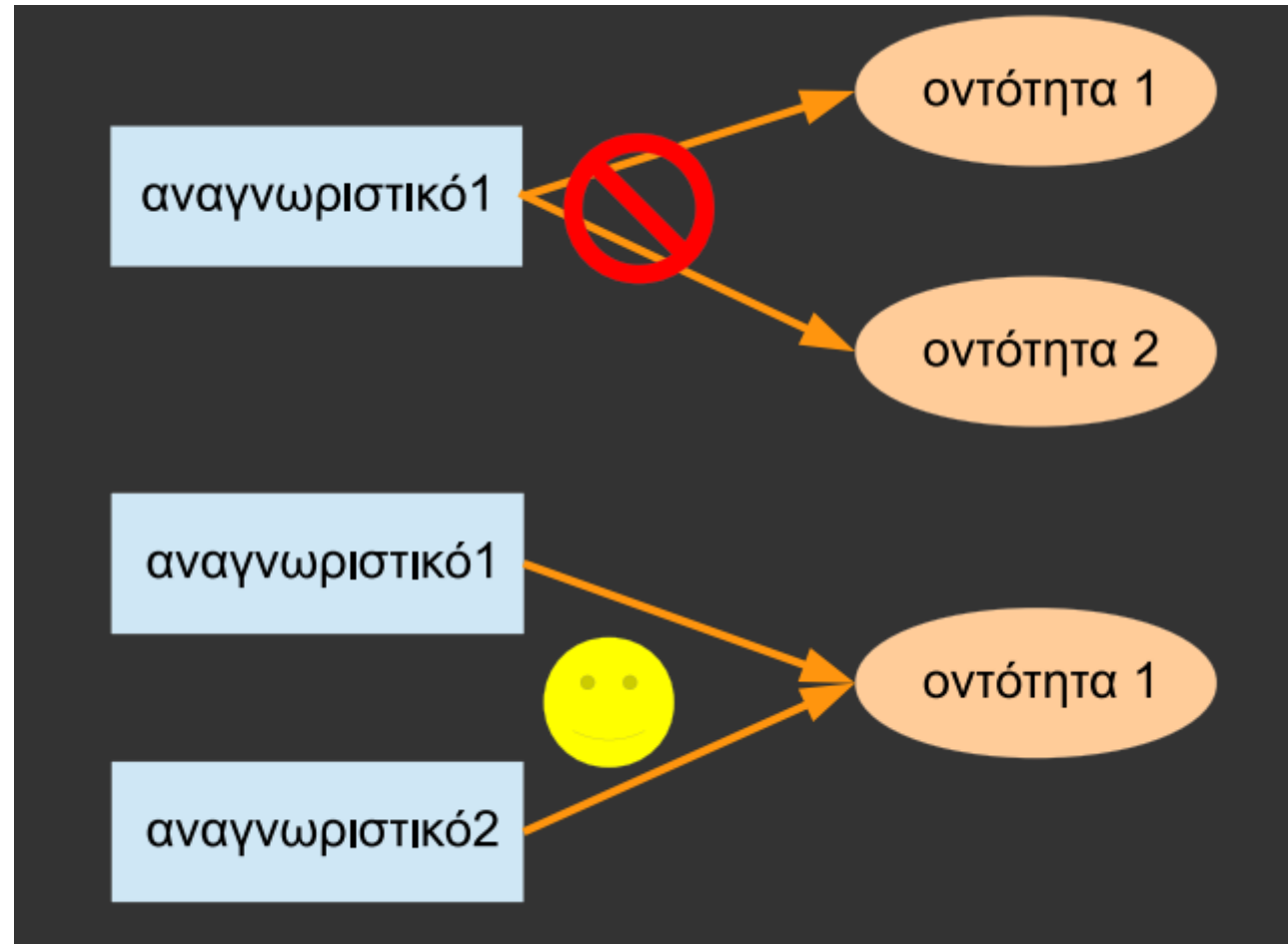
# Αναγνωριστικά

- ▶ Στον Σημασιολογικό Ιστό θέλουμε να συνδυάζουμε δεδομένα από πολλαπλές πηγές και παραγωγούς
- ▶ Συνεπώς, το ζητούμενο είναι η σφαιρική αναγνώριση των οντοτήτων
- ▶ Ένα αναγνωριστικό να υποδηλώνει μία και μόνο οντότητα, σε παγκόσμιο επίπεδο
- ▶ Στα παραδείγματά μας μέχρι τώρα χρησιμοποιούμε τοπικά αναγνωριστικά
- ▶ Όμως: ένα αναγνωριστικό όπως π.χ. `sem_web` μπορεί να χρησιμοποιείται από τρίτους για τελείως διαφορετική οντότητα!
- ▶ Πρέπει να στηριχτούμε σε πρότυπα

# Αναγνωριστικά

- ▶ Επιτρέπεται ένα αναγνωριστικό να αναφέρεται σε δύο διαφορετικές οντότητες;
  - ΟΧΙ!!! Ένα αναγνωριστικό προσδιορίζει μοναδικά μια οντότητα
- ▶ Δύο διαφορετικά αναγνωριστικά μπορούν να αναφέρονται στην ίδια οντότητα;
  - ΝΑΙ!!! Αυτό είναι απόλυτα επιτρεπτό
  - π.χ. τα
  - [http://dbpedia.org/resource/Mount\\_Olympus](http://dbpedia.org/resource/Mount_Olympus)
  - <http://sws.geonames.org/734890/>
  - αναφέρονται στην ίδια οντότητα (το βουνό Όλυμπος)
- ▶ Πρακτικά: στο δικό μας σετ δεδομένων καλό είναι να χρησιμοποιούμε μόνο ένα αναγνωριστικό για την ίδια οντότητα

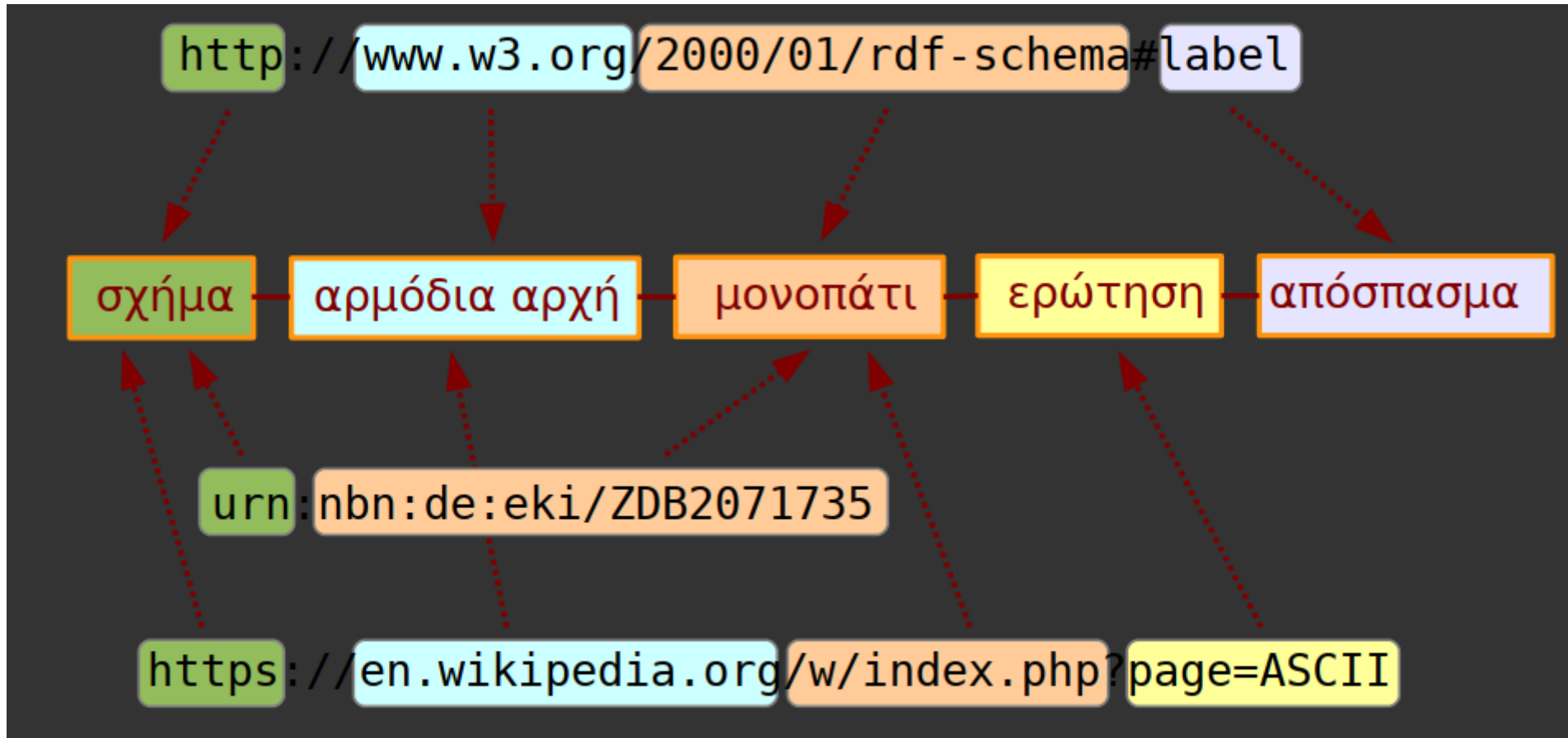
# Αναγνωριστικά



# Uniform Resource Identifiers (URIs)

- ▶ URI: ένα συμπαγές string με καλά ορισμένους κανόνες σύνταξης που αναγνωρίζει μια οποιαδήποτε οντότητα (εδώ ονομάζεται “resource”)
- ▶ Μονοσήμαντα και με παγκόσμια ισχύ
- ▶ Η οντότητα μπορεί να είναι οτιδήποτε: όχι μόνο μια πληροφοριακή πηγή (όπως ένα έγγραφο ή μια ιστοσελίδα) αλλά και άνθρωπος, πράγμα, έννοια, συναίσθημα, κλπ
- ▶ Το URI δεν είναι η οντότητα αλλά αναφέρεται στην οντότητα

# Παράδειγμα



# Κατηγορίες URIs

- ▶ Μια μεγάλη κατηγορία URIs μοιάζουν με διευθύνσεις στο Web (URLs)
- ▶ π.χ. [http://dbpedia.org/resource/Lodovico\\_Giustini](http://dbpedia.org/resource/Lodovico_Giustini)
- ▶ Μια δεύτερη κατηγορία URIs έχει τελείως διαφορετικό σχήμα:
  - `urn:uuid:f81d4fae-7dec-11d0-a765-00a0c91e6bf6`
  - `urn:nbn:de:eki/ZDB2071735`
  - `URN:ISBN:978-82-8140026-9`
- ▶ Αυτά ονομάζονται Uniform Resource Names (URNs)
- ▶ Και οι δύο τύποι απαιτούν μια εκδούσα αρχή για την μονοσήμαντη και σφαιρική ανάθεση μέρους του URI

# Uniform Resource Names (URNs)

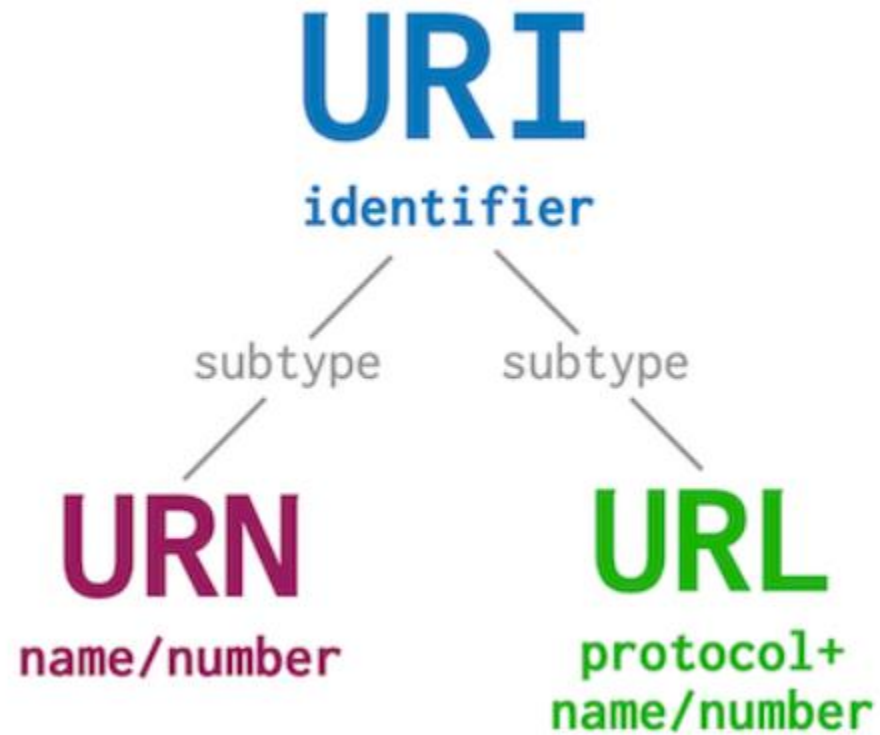
- ▶ Η αρχική ιδέα ήταν η κατασκευή ιδανικών URIs
- ▶ Με αποσύνδεση της τοποθεσίας από το αναγνωριστικό
- ▶ Σε αντίθεση με τα “μη καθαρά” `http://.. URIs` που, αν και δεν είναι υποχρεωτικό, συχνά “μπερδεύουν” την αναγνώριση μιας οντότητας με την προσπάθεια της σχετικής πληροφορίας
- ▶ Στην πράξη αποδείχτηκε ότι τα URNs είναι δύσχρηστα
- ▶ Η προσπάθεια σχετικής πληροφορίας είναι κυρίαρχη στην εποχή του Web
- ▶ Με τα URNs η προσπάθεια είναι αδύνατη:
  - πώς ξέρουμε ότι για το
  - `urn:nbn:de:eki/DNB991052625`
  - θα πάρουμε πληροφορία από το
  - `http://gso.gbv.de/DB=2.1/CMD?ACT=SRCHA&IKT=8132&TRM=DNB991052625;`
- ▶ Σήμερα, όλο και περισσότερο χρησιμοποιούνται `http URIs` στη θέση τους



# IRIs

- ▶ Σύμφωνα με τη αρχική σύνταξη των URIs [RFC3986] οι επιτρεπόμενοι χαρακτήρες σε ένα URI ανήκουν στο 7-bit ASCII (απλοί λατινικοί χαρακτήρες)
- ▶ Οι πιο κάτω χαρακτήρες είναι δεσμευμένοι και πρέπει να κωδικοποιούνται με %xx
- ▶ : / ? # [ ] @ ! \$ & ' ( ) \* + , ; =
- ▶ Επίσης τα http URIs πρέπει να κωδικοποιήσουν και τους χαρακτήρες
- ▶ < > " space { } | \ ^ `
- ▶ Τα σύγχρονα πρότυπα του Σημασιολογικού Ιστού χρησιμοποιούν τον όρο Internationalized Resource Identifiers (IRIs, [RFC3987]), όπου επιτρέπεται κάθε χαρακτήρας Unicode (εκτός των δεσμευμένων)
- ▶ Στη συνέχεια θα χρησιμοποιούμε ισοδύναμα τους όρους URI και IRI

# URI, URN, URL

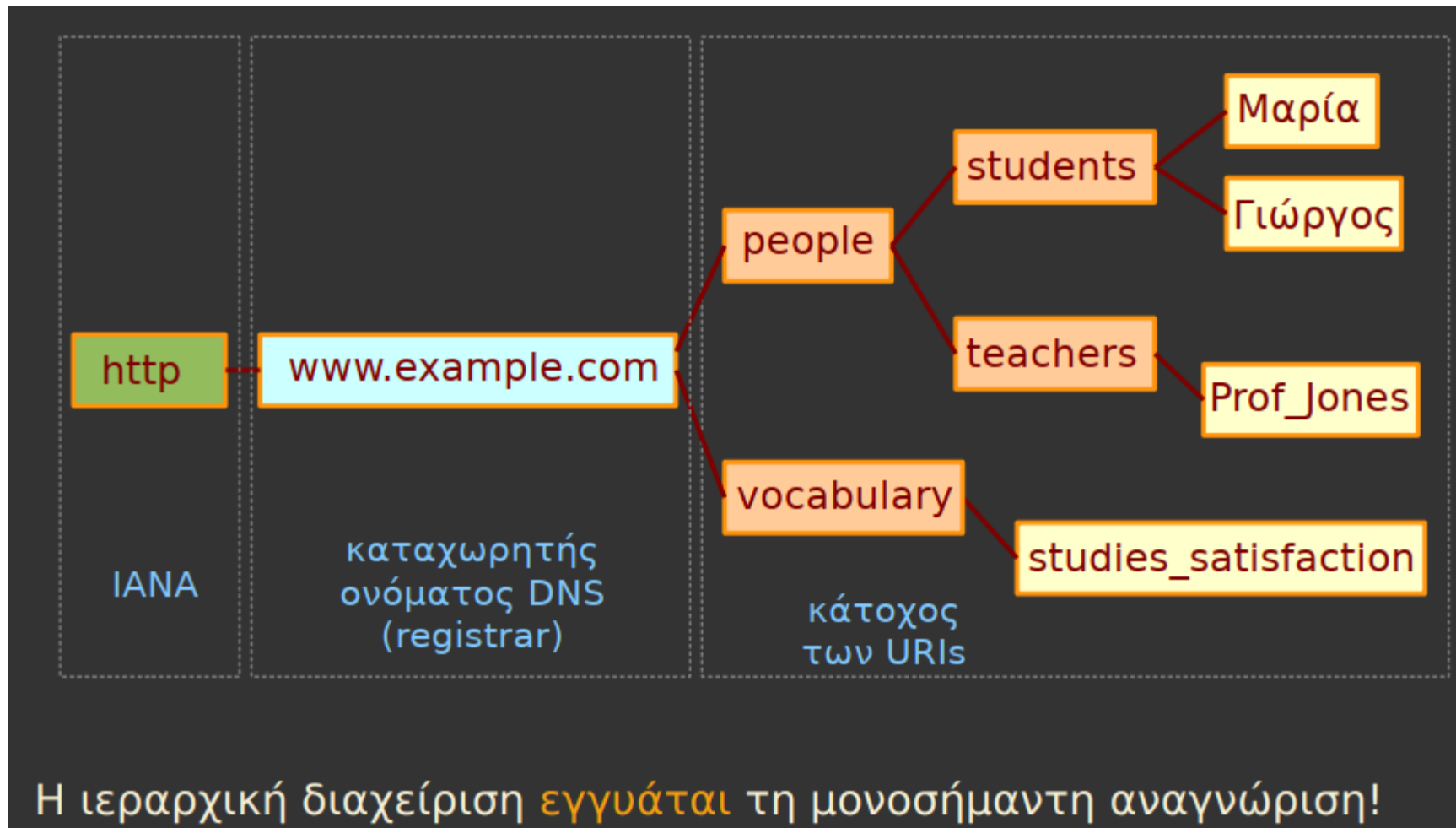


*All URLs are URIs, but not all URIs are URLs.*

# Παραδείγματα

<b>URIs</b>	<b>URNs</b>	<b>URLs</b>
A name, name and location, or both	a name or number	a name/number with location
A name, a name and address, or both	Someone's name or address	Someone's name and address
An ISBN number	An ISBN number	mailto://dan@ft.io
micah@google.com	micah@google.com	mailto://dan@ft.io
dan@ft.io	ftp.google.com	ftp://ftp.google.com

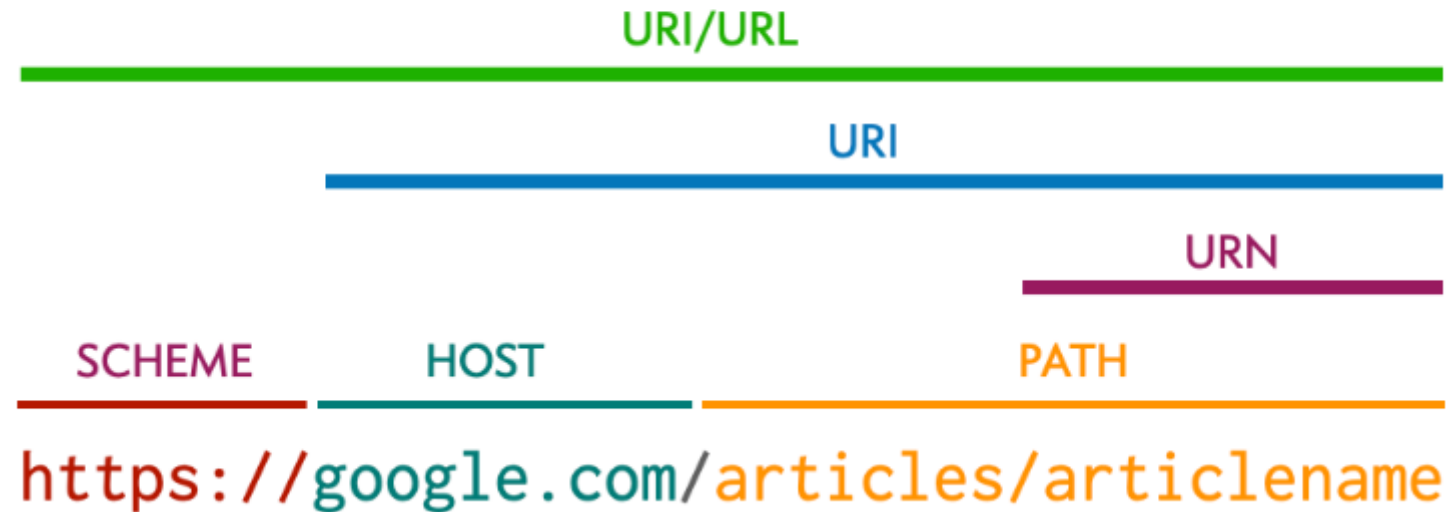
# Διαχείριση URIs/IRIs



Η ιεραρχική διαχείριση **εγγυάται** τη μονοσήμαντη αναγνώριση!

# URIs-URLs-URNs

## URL STRUCTURES



# URI, URL, or a URN?

- ▶ [www.google.com](http://www.google.com)

- ▶ [userstats.html](#)

# URI, URL, or a URN?

▶ [www.google.com](http://www.google.com)

▶ **Answer:** It's an incomplete URL because it doesn't have a protocol (although you can argue that it can be implied). As for the structure, if this were a URL it would only be the *host* portion because it lacks a *scheme* and a *path*.

▶ [userstats.html](#)

▶ **Answer:** This looks to be a resource within a URL, but since it doesn't seem to be a unique resource or prefaced by the urn: prefix, it is not a formal URN. So it is neither a URN, URL, or URI.

# Παράδειγμα: DOI

- ▶ DOI: Digital Object Identifier
- ▶ A DOI is a persistent identifier or handle used to uniquely identify various objects, standardized by the International Organization for Standardization (ISO). DOIs are an implementation of the Handle System; they also fit within the URI system (Uniform Resource Identifier). They are widely used to identify academic, professional, and government information, such as journal articles, research reports, data sets etc.
- ▶ Syntax:
- ▶ `prefix/suffix`
- ▶ E.g.
- ▶ `doi: 10.1109/TNSRE.2018.2791936`
- ▶ URL: <https://doi.org/10.1109/TNSRE.2018.2791936>
- ▶ Μας οδηγεί στο έγγραφο:
- ▶ `https://ieeexplore.ieee.org/document/8254390/`



# Χώροι ονομάτων (Namespaces)

- ▶ Κάθε URI χωρίζεται σε δύο μέρη:
  - Πρόθεμα (prefix) που ορίζει τον "χώρο διευθύνσεων" για κάθε ομάδα URIs
  - Τοπικό μέρος (local part), αναφέρεται στην οντότητα καθαυτή
- ▶ "slash (/) namespace"
  - `http://ex.com/resource/ entityA`
- ▶ "hash (#) namespace"
  - `http://ex.com/vocab# termX`
- ▶ Τα "hash" και "slash" URIs είναι ισοδύναμα ως προς τη χρήση ως αναγνωριστικά
  - Τα πράγματα αλλάζουν όταν χρησιμοποιούνται και για προσπέλαση (θα το δούμε αργότερα)

# Λεξιλόγια (Vocabularies)

- ▶ Στον Σημασιολογικό Ιστό, τα λεξιλόγια είναι ομάδες URIs, σε έναν κοινό χώρο ονομάτων, για την περιγραφή όρων σχετικών με το εκάστοτε πεδίο εφαρμογής
- ▶ Παράδειγμα: έστω (υποθετικό) λεξιλόγιο για την περιγραφή εργασιακών σχέσεων, το οποίο περιέχει τα URIs:
  - <http://ex.com/evocab#Employee>
  - <http://ex.com/evocab#Employer>
  - <http://ex.com/evocab#salary>
  - <http://ex.com/evocab#worksAt>
  - κ.ο.κ.

# ευρέως γνωστά λεξιλόγια

- ▶ Η χρήση ευρέως γνωστών λεξιλογίων στα σημασιολογικά δεδομένα
- ▶ Επιτρέπει την κατασκευή έξυπνων εφαρμογών που μπορούν να “κατανοήσουν” τη σημασία των δεδομένων
- ▶ και των σχέσεων μεταξύ δεδομένων
- ▶ Ένα γνωστό λεξιλόγιο δρα ως κοινός σημασιολογικός παρονομαστής
- ▶ Έτσι, πριν φτιάξουμε το δικό μας, πρέπει να αναζητήσουμε ήδη υπάρχοντα λεξιλόγια