

Ταυτότητα ψηφιακών αντικειμένων σε πληροφοριακά δίκτυα

ΠΑΣΧΑΛΗΣ ΡΑΠΤΗΣ¹ - ΘΕΟΔΩΡΟΣ ΚΥΡΚΟΥΔΗΣ¹¹

Περίληψη

Ο προσδιορισμός ταυτότητας σε ένα τεκμήριο (άρθρο, βιβλίο, περιοδικό) είναι χρήσιμη και γνωστή στους εκδότες, βιβλιοπώλες, βιβλιοθηκονόμους, ακαδημαϊκούς, και ερευνητές. Οι κωδικοί όπως ISBN, ISSN είναι χρήσιμοι και εξυπηρετούν όλες τις ομάδες που ασχολούνται με βιβλία και περιοδικά. Με την εμφάνιση ψηφιακών τεκμηρίων στο Διαδίκτυο έγινε επιτακτική ανάγκη ορισμού νέων κωδίκων όπως των URL & URN, PURL, SICI, DOI κλπ.

Digital object identifiers (doi) in information networks

PASCHALIS RAPTIS¹ - THEODOROS KYRKLOUDIS¹¹

Abstract

Identifiers such ISBN for books and ISSN for serials are useful and facilitate communication among booksellers, publishers, libraries and scholars. With the appearance of digital documents in a networked environ-

¹Τμήμα Πληροφορικής, ΤΕΙ Θεσσαλονίκης, Σίνδος Θεσ/νίκη, praptis@alpha.it.teithe.gr -
Department of Computer Science, TEI of Thessaloniki, e-mail: praptis@alpha.it.teithe.gr

¹¹Βιβλιοθήκη Τμήματος Ιατρικής, Δημοκρίτειο Πανεπιστήμιο Θράκης - Department of Medicine
Library, Demokritian University of Thrace

ment it has seen the emergence of new identifiers such as URL & URN, PURL, SICI, DOI etc.

Εισαγωγή

Οι γνωστοί κωδικοί όπως ISBN και ISSN είναι χρήσιμοι στις κοινότητες που ασχολούνται με την έκδοση, διακίνηση και χρήση έντυπων εκδόσεων. Οι κωδικοί αυτοί δίνουν μια ταυτότητα (identifier) μοναδική σε ένα βιβλίο ή ένα περιοδικό. Επίσης μια βιβλιογραφική αναφορά σε ένα τεκμήριο μπορεί να θεωρηθεί ως μια ταυτότητα για το τεκμήριο.

Η απόδοση ταυτότητας σε αντικείμενα είναι δύσκολη καθώς τα δεδομένα αυξάνονται με ιλιγγιώδη ρυθμούς. Η συλλογή της Βιβλιοθήκης του Κογκρέσου είναι 30 τρισεκατομμύριο τεκμήρια. Η νομική βάση Lexis-Nexis είναι περίπου 6 τρισεκατομμύριο και πρόσθεσε το δισεκατομμυριοστό (1,000,000,000) τεκμήριο το 1996, και αυξάνεται με μια ταχύτητα από περίπου 30 εκατομμύρια τεκμήρια το τρίμηνο. Στις θετικές επιστήμες εκδίδονται περίπου 650,000 τεκμήρια με μια αύξηση του 2,5% ανα έτος. Το Διαδίκτυο (Internet) περιέχει 50 εκατομμύρια μικρότερα τεκμήρια. Οι σελίδες στο Web το 1996 ήταν 76 εκατομμύρια τον Μάρτιο του 1998 ήταν 320 εκατομμύρια και στα αμέσως επόμενα χρόνια το μέγεθος του Διαδικτύου αναμένεται να αυξηθεί κατά 1000%.

Με την εμφάνιση του διαδικτύου άρχισε η έκδοση, διακίνηση και χρήση ηλεκτρονικών εκδόσεων - τεκμηρίων όπως αρχεία, βιβλία, περιοδικά, άρθρα, περιλήψεις (abstracts), φωτογραφίες, video, μουσική, δεδομένα, εργασίες πολυμέσων κλπ. Τα ηλεκτρονικά τεκμήρια είναι σε ψηφιακή μορφή και γνωστά ως ψηφιακά αντικείμενα.

Στον γρήγορα αναπτυσσόμενο χώρο της ηλεκτρονικής έκδοσης εμφανίστηκαν προβλήματα όπως η ιδιοκτησία της πληροφορίας μπορεί να αλλάζει, επίσης η ονομασία και η τοποθεσία ενός αρχείου μπορεί αλλάζει συνέχεια. Πολλές φορές οι διευθύνσεις του Διαδικτύου (τα URLs) δεν δουλεύουν επειδή οι πηγές και τα τεκμήρια στο διαδίκτυο μετακινούνται, αλλάζουν ονόματα, αλλάζουν μέθοδο πρόσβασης και για πολλές άλλες αιτίες. Όταν ένα URL αποτυγχάνει

τότε οπουδήποτε υπάρχει αυτό το URL ως αναφορά-σύνδεση καθίσταται άκυρο. Οι χρήστες του Διαδικτύου περιμένουν ότι οι πηγές και τα τεκμήρια είναι σταθερά και δεν λαμβάνουν μηνύματα όπως Error 404 File not found ή Unable to locate Server.

Τα συστήματα απόδοσης κωδίκων (ταυτοτήτων) σε ψηφιακά αντικείμενα προσπαθούν να λύσουν αυτά τα προβλήματα. Νέα συστήματα έχουν αναπτυχθεί ενώ αλλά βρίσκονται σε εξέλιξη. Ένα σύστημα κωδικοποίησης πρέπει να είναι ευέλικτο και συγχρόνως να είναι εύρηστο. Οι κωδικοί μπορεί να είναι "απλοί" πχ. ένας αριθμός ή "σύνθετοι" όπως πχ. το ISBN οπού μέσα στον κωδικό υπάρχουν πληροφορίες για την χώρα έκδοσης ή την γλωσσά.

Η απόδοση μοναδικών κωδίκων (ταυτοτήτων) σε ψηφιακά αντικείμενα είναι ζωτική και επιτρέπει:

- Ηλεκτρονικό εμπόριο και δικαίωμα συναλλαγής
- Διαχείριση πνευματικών δικαιωμάτων
- Ηλεκτρονικούς πίνακες περιεχομένων
- Παρακολούθηση παραγωγής και εσωτερική διαχείριση
- Βιβλιογραφικό έλεγχο και εντοπισμό πηγών

Η απόδοση ταυτότητας σε ψηφιακά αντικείμενα αφορά πολλά πεδία και κοινοίτητες εκτός των εκδοτών όπως: διαχείριση πνευματικών δικαιωμάτων, τεχνολογία διαδικτυου, καταλογογράφηση και βιβλιοθήκες, ηλεκτρονικό εμπόριο, πληροφορική.

URL και URN

Μια ομάδα απο κωδικούς Uniform Resource Locators (URLs) εμφανίσθηκε με την ανάπτυξη του Διαδικτυου. Τα URLs είναι χρήσιμα, δεν είναι ονόματα ώστε να καθορίζουν τα περιεχόμενα αλλά οδηγίες για το πως θα μπόρεση κάποιος να έχει πρόσβαση σε ένα αντικείμενο. Τα URLs σχεδιάστηκαν να είναι ευέλικτα, εύκολα υλοποιήσιμα, και με εύκολους τρόπους επεκτάσιμα για να κάνουν αναφορές σε υλικό στο διαδικτυο. Οι περιορισμοί των URLs στην σαφή-ακριβή αναφορά για ανάκτηση ψηφιακών αντικείμενων έφερε την ανάπτυξη ενός παραλλήλου συστήματος με το όνομα Uniform Resource Names (URN). Η σύνταξη ενός URN για ένα ψηφιακό αντικείμενο ορίζεται και περιλαμβάνει έναν κωδικό-καθιερωμένο-

όνομα (το καθιερωμένο όνομα δίδεται από ένα κεντρική γραμματεία) και ένα κωδικό για το αντικείμενο (το οποίο δίδεται από τον υπεύθυνο του καθιερωμένου ονόματος)

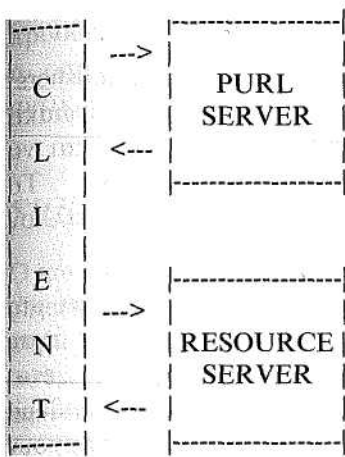
Η σύνταξη του URN δεν καθορίζει μια υπηρεσία πρόσβασης για το αντικείμενο όπως συμβαίνει με το URL. Το URN προσφέρει υπηρεσίες ανάλυσης που μεταφράζει ένα URN σε οδηγίες για πρόσβαση σε ένα αντικείμενο. Συστήματα που παρέχουν υπηρεσίες ανάλυσης ονομάζονται "αναλυτές" (resolvers) ή "βάσεις αναλυτών" (resolution databases) και μετατρέπουν το όνομα σε τοποθεσίες που υπάρχει το αντικείμενο. Ένα URN αναλύεται σε ένα σετ από URLs το οποίο παρέχει πρόσβαση σε αντιγραφο/α.....του-ζητουμένου αντικείμενου.

Προς το παρόν οι εμπορικοί περιδιαβαστες (browsers) δεν μπορούν να καταλάβουν URNs και το πως να καλέσουν ένα "αναλυτή" για να μετατρέψει ένα URN σε URLs.

PURL

Πριν μερικά χρόνια το OCLC ανέπτυξε ένα σύστημα το οποίο ονομάζεται Persistent URL (PURL). Το σύστημα αναπτύχθηκε για να λύση τα προβλήματα που προκύπτουν από μη σαφή-σταθερά URLs. Στην ουσία τα PURLs είναι HTTP URLs όπου το hostname αντικαθίσταται από τον host "PURL.ORG" και το όνομα του αρχείου είναι ένας κωδικός. Σε μια βάση δεδομένων στο OCLC ο κωδικός συνδέεται με την HTTP διεύθυνση που βρίσκεται το αντικείμενο. Ο PURL server αναζητεί τον κωδικό στην βάση, βρίσκει την πραγματική διεύθυνση και στην συνέχεια ανακατευθύνει και συνδέεται στο πραγματικό host όπου υπάρχει το αντικείμενο.

Τα PURLs δουλεύουν με τους υπάρχοντες browsers.



Παραδειγματα

To URL:

http://my.address.com/users/staff/fil_nam.txt

μπορει να συνδεθει με το PURL:

<http://purl.org/lib/fil1>

 / | \
πρωτοκολλο διευθυνση κωδικος

To URL:

<http://aaa.bbb.edu/A/B/C/D/E/f01.txt>

συνδεεται με το PURL:

<http://purl.org/TEI/12>

 / | \
πρωτοκολλο διευθυνση κωδικος

DOI

Η Ένωση Εκδοτών Αμερικής (Association of American Publishers) σχεδίασε ένα σύστημα για την απόδοση κωδίκων σε ψηφιακά αντικείμενα για να διευκολύνει την ηλεκτρονική εμπορία και να εισάγει συστήματα διαχείρισης πνευματικής ιδιοκτησίας. Το σύστημα ονομάζεται Digital Object Identifiers (DOI). (Όνομα μάλλον πολύ γενικό)

Τα DOIs δίνουν ταυτότητα σε υλικό εκδοτών. Ένα DOI είναι μια ακολουθία από χαρακτήρες. Η βασική διαδικασία είναι η εξής: α) σε κάθε εκδότη δίδεται ένας αριθμητικός κωδικός που ονομάζεται "πρόθεμα" (prefix), β) ο εκδότης δημιουργεί DOIs με βάση το πρόθεμα και γ) τους καταχωρεί στο σύστημα του DOI καταλόγου. Υπάρχουν δυο κανόνες: α) κάθε DOI πρέπει να είναι μοναδικό, β) κάθε DOI πρέπει να αρχίζει με το αριθμητικό πρόθεμα ακολουθεί μια κάθετος (/) και τέλος ακολουθεί ο κωδικός που δίνει ο εκδότης σε ένα τεκμήριο.

Ένας κωδικός DOI μπορεί να καθορίζει περιεχόμενα απλά όπως μια φωτογραφία ή περίπλοκα όπως μια εγκυκλοπαίδεια. Επίσης μπορεί να καθορίζει ένα βιβλίο, ένα κεφαλαίο από ένα βιβλίο ή μια φωτογραφία σε ένα κεφαλαίο, περιοδικά, ηλεκτρονικές πληροφορίες, μουσικούς δίσκους, video κλπ.

Παραδείγματα

DOI:

10.10001/45678

/ \

κωδ. εκδότη κωδ. τεκμηρίου

DOI και το σχετικό URL:

10.1002/0002-8231(199803)48:2<133:ESOPAN>2.3.TX;2-Q
<http://journals.wiley.com/0002-8231/abs/v48n2p133.html>

Ο κωδικός τεκμηρίου μπορεί να είναι ένας αριθμός ή ένας γνωστός κωδικός όπως ISBN, ISSN κλπ.

Το σύστημα DOI συνδέει χρήστες, εκδότες, και πληροφορίες, αποτελείται από κωδικούς DOI και μια βάση οπου τοποθετούνται αυτά τα DOI. Η βάση ονομάζεται Κατάλογος DOI.

Με έναν browser ο χρήστης αναζητά το κατάλογο-DOI. Το σύστημα προωθεί την αναζήτηση στην βάση του εκδότη ή σε έναν Web server, ο οποίος επιστρέφει ένα αντικείμενο ή πληροφορίες για το αντικείμενο στον χρήστη.

SICI

Ο κωδικός SICI (Serial Item and Contribution Identifier) ορίζει τεύχη ("items") ενός περιοδικού και άρθρα ("contributions") σε ένα τεύχος περιοδικού. Η ανανεωμένη έκδοση του SICI το 1997 έχει βελτιωθεί ώστε να είναι δυνατόν να ορισθεί: το μέσον έκδοσης (medium of publication) και τα συστατικά του (components).

Ο κωδικός SICI αποτελείται από τρία τμήματα:

α) τμήμα του τεύχους (item),

Παράδειγμα

1234-5679(19960221)1:1

(Σημαίνει: ISSN 1234-5679, εκδόθηκε το 21 Φεβ. 1996, τόμος 1, τεύχος 1)

β) τμήμα του άρθρου (contribution),

Παράδειγμα

<123:ABCDEF>

(Σημαίνει: Τοποθεσία 123-135, κωδικός τίτλου ABCDEF) γ) τμήμα ελέγχου.

Παράδειγμα

2.0.TX;2-A

(Σημαίνει: CSI2, κείμενο, αριθμός έκδοσης 2, χαρακτήρας ελέγχου A)

Στην νέα έκδοση εισήχθη ο CSI (Code Structure Identifier) που έχει μια από τις παρακάτω τρεις τιμές. Έτσι ο SICI μπορεί να πάρει τρεις μορφές:

CSI 1 (μορφή 1) ορίζει το τεύχος περιοδικού

Παράδειγμα

1234-5678(19960221)1:2:3<>1.0.TX;2-P

CSI 2 (μορφή 2) ορίζει ένα συστατικό του τεύχους

Παράδειγμα

1234-5678(19960221)1:2:3<123:ABCDEF>2.0.TX;2-A

CSI 3 (μορφή 3) ορίζει ένα συστατικό αλλά μπορεί να ενσωματώσει και άλλους κωδικούς

Παράδειγμα

1234-5678(1996)<:::INS-023456>3.0.CO;2-#

(Η CSI 3 μορφή κωδικοποίησης είναι ευέλικτη όπως και το ΡΠ. Ο SICI και ο ΡΠ είναι συμβατοί και οι δυο κωδικοί μπορούν να συνυπάρχουν αρμονικά και το ένα μπορεί να συμπληρώνει τον άλλον)

BICI

Ο κωδικός Book Item and Component Identifier (BICI) είναι μια ακολουθία χαρακτήρων μεταβλητού μήκους που δίνει ταυτότητα σε βιβλία (πχ. τόμους) και σε συστατικά (πχ. κεφάλαια) ενός βιβλίου. Ο BICI μπορεί να εφαρμοσθεί σε μη-περιοδικές εκδόσεις που πλήρη δυο συνθήκες: 1) η έκδοση έχει ένα ISBN, και 2) η έκδοση είναι σελιδοποιημένη ή έχει σαφείς επιμέρους τίτλους ή ο εκδότης έχει δώσει ταυτότητα σε κάθε συστατικό της έκδοσης.

Δεν υπάρχει περιορισμός για το μέσον στο οποίο η έκδοση έχει εμφανισθεί.

Ο κωδικός BICI μπορεί να δοθεί σε:

- α) μέρος (part), κεφαλαίο ή τμήμα σε κεφαλαίο (μη-κείμενο όπως φωτογραφίες, λίστες, ταμπέλες κλπ ή με κείμενο όπως εισαγωγή, πρόλογος, επίλογος, ευρετήριο, βιβλιογραφία κλπ)
- β) μια καταχώρηση ή άρθρο σε έναν κατάλογο ή εγκυκλοπαίδεια
- γ) ένα σχήμα, χάρτη, σχέδιο, φωτογραφία κλπ
- δ) ένθετο σε μια έκδοση

Ο κωδικός BICI αποτελείται από τρία τμήματα:

- α) τμήμα του τεκμηρίου,

Παράδειγμα

185715780X(1996)1:1

(Σημαίνει: ISBN 185715780X, εκδόθηκε το 1996, μέρος 1, τόμος 1)

- β) τμήμα του συστατικού,

Παράδειγμα

<123-135;TLSPR;7>

(Σημαίνει: Σελίδες 123-135, κωδικός τίτλου TLSPR, κεφαλαίο 7)

- γ) τμήμα ελέγχου.

Παράδειγμα

B.T.TX;1-X

(Σημαίνει: Κωδικός δομής τύπος B, κείμενο, τυπωμένη έκδοση, σταθερότυπο BICI 1, χαρακτήρας ελέγχου X)

(Ο κωδικός δομής μπορεί να πάρει τις τιμές A (βιβλίο), B (μέρος βιβλίου), C (μέρος βιβλίου που ορίζεται από τον εκδότη))

Παραδείγματα

1) Ο BICI ορίζει ένα βιβλίο:

185715780X(1996)1:K>A.X.TX;1-X

2) Ο BICI ορίζει ένα μέρος (συστατικό) ενός βιβλίου

185715780X(1996)<123-135;TLSPR;7>B.T.TX.1-X

3) Ο BICI ορίζει ένα μέρος (συστατικό) ενός βιβλίου με κωδικό που δόθηκε από τον εκδότη (G123456)

185715780X(1996)<1-10;ITB;l;G123456>C.T.TX;1-X

ή

185715780X(1996)<;;G123456>C.T.TX;1-X

PII

Ο Publisher Item Identifier (PII) είναι ένα σύστημα απόδοσης απλού κωδικού το οποίο σχεδιάστηκε από μια μικρή ομάδα εκδοτών επιστημονικών-εκδόσεων (όπως η Αμερικανική Ένωση Χημικών, το Αμερικανικό Ίδρυμα Φυσικής, το IEEE κλπ) Ο σκοπός ήταν να βρεθεί ένα σύστημα απόδοσης κωδικού καθώς οι επιχειρήσεις έπρεπε να περάσουν στην διακίνηση τεκμηρίων στον ηλεκτρονικό κόσμο.

Το PII κωδικός είναι μια ακολουθία από 17 χαρακτήρες (ο 17ος χαρακτήρας είναι χαρακτήρας ελέγχου). Για να μπορέσει ο PII κωδικός να είναι μοναδικός δημιουργείται από το ISBN ή το ISSN. Μπορεί να επεκταθεί και σε άλλες τύπους τεκμηρίων. Το PII εισήχθη το 1996 και το υιοθέτησαν πολλοί εκδότες με εκδόσεις στις 3ετικες-επιστημες.

Το σύστημα PII μπορεί να χρησιμοποιηθεί δωρεάν από οποιονδήποτε εκδότη ή άλλον ενδιαφερόμενο που επιθυμεί να το εφαρμόσει. Παραδείγματα

ΡΠ: S 01653806 96 004038

| ISSN έτος Τεκμήριο |
Serial Χαρ. Ελέγχου

ΡΠ: B 0165380696 00 4038

| ISBN Τεκμήριο |
Book Χαρ. Ελέγχου

ISWC

Ο κωδικός ISWC (International Standard Work Code) είναι ένας αριθμός που χρησιμοποιείται για να ορίσει ένα μουσικό έργο ή μια σύνθεση. Ο κωδικός έχει δυνατότητα επέκτασης και σε άλλες μορφές πληροφορίας. Δημιουργήθηκε από συνθέτες, μουσικούς οίκους, και κοινότητες προστασίας πνευματικής ιδιοκτησίας για να καλύτερη προστασία και διαχείριση της μουσικής σε παγκοσμία κλίμακα στην ψηφιακή μας εποχή. Ο κωδικός ISWC είναι ένας "απλός" κωδικός και μπορεί να υπάρχουν 100 εκατομμύρια διαφορετικοί κωδικοί. Ο ISWC μπορεί να χρησιμοποιηθεί σε συνδυασμό με ήδη υπάρχοντες κωδικούς όπως ISBN, ISRC κλπ.

Επίλογος

Τα δίκτυα H/Y είναι ένα νέο περιβάλλον για πρόσβαση και παροχή πληροφοριών και η απόδοση ταυτότητας σε τεκμήρια είναι ζωτική σε ένα τέτοιο περιβάλλον.

Νέα συστήματα απόδοσης κωδίκων σε ψηφιακά αντικείμενα εμφανίζονται ενώ αλλά καταργούνται.

Η απόδοση κωδίκων σε ψηφιακά αντικείμενα φαίνεται να είναι ένας τρόπος για προστασία και εξυπηρέτηση των χρηστών και των εκδοτών.

Υπάρχουν πολλές κοινότητες που ευλόγα ενδιαφέρονται να προσεγγίσουν από πολλές κατευθύνσεις στο πρόβλημα καθιέρωσης

νέων συστημάτων κωδικοποίησης των ψηφιακών αντικείμενων από τον κόσμο της μουσικής, από τους παραδοσιακούς εκδότες, και από το Διαδίκτυο.

Η δημιουργία νέων κωδίκων πρέπει να γίνει με συνεργασία μεταξύ της κοινότητας που χρησιμοποιούν παραδοσιακά σταθερότυπα και της ταχέως αναπτυσσόμενης κοινότητας του Διαδικτύου.