

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ

ΤΜΗΜΑ ΨΗΦΙΑΚΩΝ ΣΥΣΤΗΜΑΤΩΝ



Πρόγραμμα Μεταπτυχιακών Σπουδών
**Τεχνοοικονομική Διοίκηση & Ασφάλεια Ψηφιακών
Συστημάτων**

Κατεύθυνση: Ασφάλεια Ψηφιακών Συστημάτων

Διπλωματική Εργασία

Μελέτη των Μεταδεδομένων σε αρχεία του Microsoft Office X

Χριστόδουλος Πάσχος

Επιβλέπων: Επίκουρος Καθηγητής Χρήστος Ξενάκης

Πειραιάς, Απρίλιος 2014

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ

[Κενή Σελίδα]

Περίληψη

Στην παρούσα διπλωματική εργασία για την απόκτηση Μεταπτυχιακού Διπλώματος Ειδίκευσης, μελέτησα τα μεταδεδομένα σε αρχεία του Microsoft Office X. Αρχικά, έγινε ο προσδιορισμός της έννοιας των μεταδεδομένων και παρουσιάστηκε η χρησιμότητά τους στην εγκληματολογική έρευνα. Στη συνέχεια, αναλύθηκαν οι τύποι των μεταδεδομένων σε προγράμματα της σουίτας του Microsoft Office και το είδος των πληροφοριών που αποκαλύπτουν και παρέχονται οδηγίες πρόσβασης σε αυτά ή στιγμιότυπα από τα ίδια τα προγράμματα. Έπειτα, αναφέρεται εκτενώς το ενσωματωμένο εργαλείο που παρέχει η Microsoft για την ανάκτηση καθώς και αφαίρεση μεταδεδομένων με παραδείγματα και οδηγίες. Ακολούθως, παρέθεσα μερικά από τα δημοφιλέστερα εμπορικά εργαλεία προβολής και απομάκρυνσης μεταδεδομένων και κάποια εργαλεία με άδεια ελεύθερης χρήσης. Τέλος, μελέτησα τις περιπτώσεις ανάκτησης μεταδεδομένων από δείγματα αρχείων του MS Word, παλαιού και νέου τύπου, με χρήση εργαλείων ελεύθερης χρήσης, και συγκρίνουμε τα αποτελέσματα.

Λέξεις κλειδιά: μεταδεδομένα, Microsoft Office, ψηφιακά πειστήρια, προστασία απορρήτου

Abstract

In the present thesis for the acquisition of a Master of Science Degree, I study the metadata in Microsoft Office X files. Initially, the meaning of metadata is defined and I present their usefulness in forensic analysis and privacy protection. Subsequently, I analyze the types of metadata in programs of the Microsoft Office X suite and the kind of information they reveal, while access instructions to them or screenshots of the programs themselves are provided. Microsoft's incorporated tool for the acquisition and removal of metadata is extensively discussed, providing examples and instructions. In addition, I mention some of the most popular commercial tools for viewing and clearing metadata and some open-source or freeware tools. Finally, I study the cases of metadata extraction from MS Word new and old file-type samples, using open-source tools and then I compare the results.

Keywords: metadata, Microsoft Office, digital forensics, privacy protection

Ευχαριστίες

Η παρούσα διπλωματική εργασία εκπονήθηκε στα πλαίσια του Μεταπτυχιακού Προγράμματος Σπουδών «Τεχνοοικονομική Διοίκηση & Ασφάλεια Ψηφιακών Συστημάτων» του Τμήματος Ψηφιακών Συστημάτων, του Πανεπιστημίου Πειραιώς, υπό την επίβλεψη του Επίκουρου Καθηγητή κ. Χρήστου Ξενάκη. Θα ήθελα να ευχαριστήσω θερμά τον κύριο Χρήστο Ξενάκη, για την ευκαιρία που μου έδωσε να ασχοληθώ με ένα τόσο ενδιαφέρον αντικείμενο καθώς και για την αμέριστη συμπαράστασή του καθ' όλη την διάρκεια εκπόνησης.

Επιπλέον θα ήθελα να ευχαριστήσω τον κ. Χριστόφορο Νταντογιάν – επιστημονικό συνεργάτη του κ. Ξενάκη - για την συνεργασία, τις πολύτιμες υποδείξεις και τη συμβολή του στην ολοκλήρωση της παρούσας εργασίας.

Τέλος, θα ήθελα να ευχαριστήσω τους συναδέλφους, συμφοιτητές και φίλους, Χαρακίδα Αντώνη και Χρήστο Αλέξη, για το ωραίο ταξίδι της γνώσης που πραγματοποιήσαμε ως συνοδοιπόροι.

Πάνω από όλα, ένα μεγάλο ευχαριστώ πηγαίνει στη γυναίκα μου Στέλλα, για το υπέροχο δώρο που έφερε στη ζωή μας αλλά και για τη συνεχή συμπαράστασή και υποστήριξη που μου δίνει.

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ

[Κενή Σελίδα]

Πίνακας Περιεχομένων

1	Εισαγωγή	9
2	Μεταδεδομένα	10
3	Μεταδεδομένα στο Microsoft Office X.....	12
4	Office Open XML format	12
4.1	Διάρθρωση εγγράφου OOXML	13
4.2	Πλεονεκτήματα των μορφών Open XML	14
4.3	Επεκτάσεις ονόματος αρχείου XML	15
4.3.1	Word.....	15
4.3.2	Excel.....	16
4.3.3	PowerPoint	16
5	Εργαλεία Εύρεσης, Ανάλυσης ή Αφαίρεσης Μεταδεδομένων.....	17
5.1	Κατάργηση κρυφών δεδομένων και προσωπικών πληροφοριών με επιθεώρηση εγγράφων.....	17
5.2	Κατάργηση κρυφών δεδομένων και προσωπικών πληροφοριών με έλεγχο βιβλίων εργασίας.....	19
5.3	Κατάργηση κρυφών δεδομένων και προσωπικών πληροφοριών με έλεγχο παρουσιάσεων	21
5.4	Επισκόπηση βασικών Ιδιοτήτων Ελέγχου Εγγράφου	22
5.5	Τι δεν ανιχνεύει ο Έλεγχος εγγράφου;	23
5.5.1	Κρυφό ή Αόρατο Περιεχόμενο	23
5.5.2	Δεδομένα Προσωρινής Αποθήκευσης	24
5.5.3	Εξωτερικές συνδέσεις	25
5.5.4	Προστατευμένα και περιορισμένα αρχεία	25
5.5.5	VBA και ActiveX.....	26
5.5.6	Άλλες πηγές μεταδεδομένων.....	26
5.6	Επισκόπηση σύνθετων Ιδιοτήτων Ελέγχου Εγγράφου.....	27
5.7	Μικρογραφίες.....	29
5.8	Χρονοσφραγίδες.....	30
5.9	Παρακολούθηση Αλλαγών.....	30
5.10	Αποστολή για αναθεώρηση μέσω των εφαρμογών του MS Office	31
6	Άνοιγμα αρχείου .docx ως XML	32
6.1	Διάρθρωση OOXML	32
6.1.1	rels	33
6.1.2	customXml	33
6.1.3	docProps	34
6.1.4	word.....	37

6.1.5	Content_Types	38
7	Μεταδεδομένα σε Παλαιότερες Εκδόσεις του Office	39
8	Μεταδεδομένα σε Αρχεία του OpenOffice.org	43
9	Εμπορικά εργαλεία	44
10	Ελεύθερης χρήσης ή Ανοιχτού κώδικα εργαλεία	47
11	Μελέτη Περίπτωσης: Εξαγωγή Δεδομένων από τα έγγραφα του MS Word ...	48
11.1	Αυτοματοποιημένα με χρήση εργαλείου ελεύθερης χρήσης.....	48
11.1.1	Μεταδεδομένα αρχείου .docx με το εργαλείο MetaExtractor	48
11.2	Metadata Touch	54
12	Σύνοψη.....	55

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ

Πίνακας Εικόνων

Εικόνα 1 Έναρξη του στοιχείου "Έλεγχος Εγγράφου" στο Office 2007.....	17
Εικόνα 2 Έναρξη του στοιχείου "Έλεγχος Εγγράφου" στο Office 2010.....	17
Εικόνα 3 Πλαίσιο διαλόγου του Ελέγχου Εγγράφου στο MS Word 2007	18
Εικόνα 4 Προβολή των Βασικών Ιδιοτήτων Ελέγχου Εγγράφου.....	23
Εικόνα 5 Προβολή Βασικών Ιδιοτήτων Ελέγχου Εγγράφου	23
Εικόνα 6 Προβολή των Πληροφοριών Σύνοψης του Εγγράφου	28
Εικόνα 7 Προβολή Πληροφοριών Στατιστικών Στοιχείων του Εγγράφου.....	29
Εικόνα 8 Χειροκίνητη Επιλογή της Αποθήκευσης Μικρογραφίας	29
Εικόνα 9 Η Μικρογραφία ενός αρχείου .docx	30
Εικόνα 10 Ενεργοποίηση της Παρακολούθησης Αλλαγών στο Office 2007	30
Εικόνα 11 Περιεχόμενα του Εγγράφου .docx κατά το άνοιγμα ως XML	32
Εικόνα 12 Εντοπισμός μοναδικών αναγνωριστικών στο άνοιγμα ως XML.....	33
Εικόνα 13 XML κώδικας του αρχείου itemProps1.xml.....	33
Εικόνα 14 Περιεχόμενα του φακέλου docProps	34
Εικόνα 15 XML κώδικας του αρχείου custom.xml	34
Εικόνα 16 XML κώδικας του αρχείου core.xml.....	35
Εικόνα 17 XML κώδικας του αρχείου app.xml.....	36
Εικόνα 18 Πολυμέσα αρχείου word κατά το άνοιγμα ως XML	37
Εικόνα 19 Ανάκτηση Κειμένου από Οποιοδήποτε Αρχείο.....	39
Εικόνα 20 Πλαίσιο διαλόγου μετά την ανάκτηση αρχείου.....	39
Εικόνα 21 Δείγμα περιεχομένων ανακτηθέντος αρχείου.....	40
Εικόνα 22 Απομάκρυνση Προσωπικών Πληροφοριών από Ιδιότητες Αρχείου.....	41
Εικόνα 23 Αποστολή Εγγράφου για Αναθεώρηση	42
Εικόνα 24 Δείγμα μεταδεδομένων σε OpenOffice.org Writer έγγραφο.....	43
Εικόνα 25 Μεταδεδομένα που εντόπισε το MetaExtractor	48
Εικόνα 26 Εξαγωγή και αποθήκευση των Μεταδεδομένων από το MetaExtractor.....	49
Εικόνα 27 Εισαγωγή μεταδεδομένων από το MetaExtractor στο MS Excel 2007.....	49
Εικόνα 28 Επιλογή του τύπου δεδομένων στο αρχείο που θέλουμε να εισάγουμε	50
Εικόνα 29 Ορισμός ορισμένης διαφορετικών τιμών εισαχθέντων δεδομένων.....	50
Εικόνα 30 Παρουσίαση των μεταδεδομένων από το MetaExtractor στο Excel	53
Εικόνα 31 Παρουσίαση των μεταδεδομένων από το Metadata Touch.....	54

1 Εισαγωγή

Οι περισσότερες εφαρμογές λογισμικού αποθηκεύουν πληροφορία (γνωστή ως μεταδεδομένα) στα αρχεία που χρησιμοποιούν προκειμένου να παρέχουν και να διατηρούν ένα ιστορικό του αρχείου, να προωθούν τη συνεργασία μέσα σε ένα οργανισμό και να διατηρούν όλη την πληροφορία που σχετίζεται με το αρχείο σε μια κεντρική τοποθεσία. Πολλή από αυτή την πληροφορία αποθηκεύεται ως μέρος ενός χαρακτηριστικού του προϊόντος ή ως ρύθμιση ιδιότητας για να βοηθήσει στην συνεργασία. Παρόλα αυτά μερικές φορές δεδομένα αποθηκεύονται εν αγνοία του χρήστη του τι δεδομένα και πού αποθηκεύονται. Ακόμα και φαινομενικά αβλαβή χαρακτηριστικά και ρυθμίσεις σε ένα προϊόν μπορεί να αποθηκεύσει πληροφορίες που παρέχουν στοιχεία για ένα άτομο ή την εταιρία του σε αδιάκριτα βλέμματα.

Η εργασία αυτή βασίζεται κυρίως στα χαρακτηριστικά και τις λειτουργίες των Microsoft Office 2007 και 2010, καθώς είναι τα ευρέως χρησιμοποιούμενα αυτή τη στιγμή. Παρόλα αυτά, μπορεί να γίνει αναγωγή και εφαρμογή των γνώσεων που αποκομίζει ο χρήστης και σε παλαιότερες ή νεότερες εκδόσεις, αφού ο πυρήνας, η νοοτροπία και ο τρόπος ανάλυσης παραμένει παρεμφερής.

2 Μεταδεδομένα

Τα μεταδεδομένα (Metadata) είναι μια δομημένη περιγραφή αντικειμένων, που περιέχει συγκεκριμένες ιδιότητες χρήσιμες για το χρήστη καθώς και για το πρόγραμμα με το οποίο ένα έγγραφο δημιουργήθηκε. Απλοϊκά, τα μεταδεδομένα είναι δεδομένα που περιγράφουν άλλα δεδομένα.

Ένα μοναδικό χαρακτηριστικό των ηλεκτρονικών εγγραφών είναι ότι περιλαμβάνουν κρυμμένα μεταδεδομένα που περιέχουν, μεταξύ άλλων, εκτενή πληροφορία σχετικά με τη δημιουργία ενός αρχείου, την πιθανή διαγραφή ή εκτύπωσή του στο παρελθόν, τη διαδρομή όπου αποθηκεύεται σε έναν υπολογιστή ή δίκτυο, τον ίδιο τον υπολογιστή, τα άτομα που το δημιούργησαν ή επεξεργάστηκαν και πρόσθετες ιδιότητες του.

Τα e-mails περιλαμβάνουν μεταδεδομένα που υποδεικνύουν πληροφορία σχετικά με το βιβλίο διευθύνσεων του αποστολέα, την ημερομηνία που στάλθηκε, παραλήφθηκε, απαντήθηκε, προωθήθηκε, σε ποιον στάλθηκαν αντίγραφα και την παρουσία συνημμένων.

Το λογισμικό Adobe Acrobat δημιουργεί λεπτομερή πληροφορία διαδρομής μεταδεδομένων που μπορεί επίσης να παρέχει ερευνητική πληροφορία για αρχεία που δημιουργούνται ή αποθηκεύονται σε αυτά τα προγράμματα.

Τα μεταδεδομένα που είναι ενσωματωμένα σε έγγραφα HTML επιτρέπουν μια μηχανή αναζήτησης να καταλογοποιήσει τα θέματα που θίγονται στη συγκεκριμένη ιστοσελίδα, αυξάνοντας την αποτελεσματικότητα της αναζήτησης και πιθανόν την επισκεψιμότητα της ιστοσελίδας.

Τα μεταδεδομένα εφαρμογής (application metadata) δημιουργούνται αυτόματα από το λογισμικό εφαρμογής και ενσωματώνονται σε κάθε αρχείο που δημιουργείται ή υφίσταται επεξεργασία από αυτό το λογισμικό. Τα λειτουργικά συστήματα που ελέγχουν προσωπικούς υπολογιστές, εξυπηρετητές (servers) και συστήματα επικοινωνίας δημιουργούν μεταδεδομένα συστήματος (system metadata), τα οποία αποδίδουν πεδία πίνακα καταχώρησης αρχείων (όνομα αρχείου, δημιουργία, μέγεθος και χρήση) σε όλα τα αρχεία που αποθηκεύονται στο σύστημα ώστε το λειτουργικό σύστημα να τα ταυτοποιεί και εντοπίζει για μελλοντική χρήση. Τα μεταδεδομένα συστήματος βρίσκονται στο μητρώο (registry) συστήματος του υπολογιστή ή server που χρησιμοποιήθηκε για να προσπελάσει και να αποθηκεύσει το αρχείο.

Τα μεταδεδομένα αποκαλούνται το «ηλεκτρονικό ισοδύναμο του DNA» και έχουν την ικανότητα να ρίξουν φως στην προέλευση, στο περιεχόμενο, στην αυθεντικότητα και στο διαμοιρασμό ηλεκτρονικών στοιχείων.

Αν και είναι πολύ χρήσιμα για τη συνεργασία μεταξύ πολλών ανθρώπων, ένα πρόβλημα είναι ότι πολλοί χρήστες δεν είναι καλά ενημερωμένοι για το ποια πληροφορία αποθηκεύεται στα έγγραφά τους όταν τα κοινοποιούν και τα διαμοιράζονται. Αν δεν αφαιρεθούν στοχευμένα από το χρήστη, τα μεταδεδομένα σε ένα έγγραφο θα αποθηκευτούν ταυτόχρονα με κάθε κείμενο που δημιουργεί ο χρήστης. Επομένως, αυτός ο τύπος δεδομένων μπορεί να αποκαλύψει ακούσια

ευαίσθητες πληροφορίες, αλλά και να χρησιμεύσει στην ψηφιακή εξέταση υπολογιστών.

Τα μεταδεδομένα μπορούν να βοηθήσουν στην ταυτοποίηση ανθρωπίνων ή συστημικών ενεργειών σε πληροφοριακά συστήματα και μπορούν να χρησιμοποιηθούν για την έρευνα και εξακρίβωση απάτης, κατάχρησης, λαθών ή αποτυχιών συστήματος και να βοηθήσουν στην στοιχειοθέτηση κατηγορητηρίων, στη χρονολόγηση και εκτίμηση του εύρους γνώσης, ή ως τεκμήριο αθωότητας, που αποτελούν υποθέσεις εγκληματικών ή πολιτικών δικαστικών αγωγών¹. Το δικαστικό σύστημα συνολικά αναγνωρίζει τη χρησιμότητα των μεταδεδομένων και την βιωσιμότητα και αποδοχή τους ως στοιχεία.

Τα μεταδεδομένα που περιέχονται σε έγγραφα μπορεί να είναι σημαντική πηγή πληροφορίας για τους ερευνητές. Παρόλα αυτά η εγκληματολογική ανάλυση υπολογιστών και των δεδομένων που περιέχουν μπορεί να είναι μια εξαιρετικά δύσκολη και χρονοβόρα διαδικασία.

¹ <http://precisioncomputerinvestigations.wordpress.com/2010/04/14/how-computer-forensics-solved-the-btk-killer-case/>

3 Μεταδεδομένα στο Microsoft Office X

Αρχεία μεταδεδομένων εφαρμογών και συστήματος δημιουργούνται και ενημερώνονται για αρχεία MS Word, Excel και PowerPoint, κάθε φορά που ένα αρχείο δημιουργείται, ανοίγεται, ή χρησιμοποιείται

Τα μεταδεδομένα ενός εγγράφου, περιλαμβάνουν λεπτομέρειες και πληροφορίες σχετικά με το έγγραφο που τηρούνται/αποθηκεύονται αυτόματα όπως²:

- το όνομα και τα αρχικά του ατόμου που αποθήκευσε το έγγραφο για τελευταία φορά,
- τα ονόματα των προηγούμενων συγγραφέων του εγγράφου,
- την ονομασία της εταιρίας ή του οργανισμού,
- το όνομα του υπολογιστή,
- την ημερομηνία δημιουργίας του εγγράφου,
- το όνομα του εξυπηρετητή δικτύου ή του σκληρού δίσκου όπου αποθηκεύτηκε το έγγραφο,
- αναθεωρήσεις του εγγράφου,
- εκδόσεις του εγγράφου (versions),
- πληροφορίες του προτύπου (template),
- εξατομικευμένες προβολές,
- σχόλια,
- περίληψη,
- μη-ορατά τμήματα ενσωματωμένων αντικειμένων OLE
- άλλες ιδιότητες του αρχείου

Αν έχουν χρησιμοποιηθεί ειδικές δυνατότητες, το έγγραφο μπορεί επίσης να περιλαμβάνει πρόσθετα είδη προσωπικών στοιχείων ταυτότητας (PII)³, όπως:

- κεφαλίδες ηλεκτρονικού ταχυδρομείου,
 - πληροφορίες αποστολής για αναθεώρηση,
 - δελτία δρομολόγησης (routing slips),
 - διαδρομές εκτυπωτή και
 - πληροφορίες διαδρομής αρχείου για τη δημοσίευση ιστοσελίδων.
-
- Κεφαλίδες, υποσέλιδα και υδατογραφήματα

4 Office Open XML format

Το νέο format αρχείων εμφανίστηκε στη σουίτα εφαρμογών Microsoft Office 2007. Η προτυποποίηση είχε ξεκινήσει το 2006 από την ECMA⁴ μέσω της τεχνικής

² <http://support.microsoft.com/kb/223396/en-us>

³ <http://office.microsoft.com/en-us/word-help/inspect-documents-for-hidden-data-and-personal-information-HA010074435.aspx#BM1>

⁴ <http://www.ecma-international.org/publications/standards/Ecma-376.htm>

επιτροπής 45 (TC45) την οποία αποτελούσαν εκπρόσωποι των Apple, Barclays Capital, BP, The British Library, Essilor, Intel, Microsoft, NextPage, Novell, Statoil, Toshiba, η US Library of Congress⁵.

Ακολούθησαν τρεις ακόμα τροποποιήσεις το 2008, 2011 και τελευταία το 2012, η οποία, ταυτίστηκε τεχνικά με το πρότυπο ISO/IEC 29500⁶.

Οι νέες μορφές αρχείων χρησιμοποιούν την τεχνολογία ZIP για να συμπιέζουν τα αρχεία αυτόματα. Χωρίς να χρειάζεται άλλο λογισμικό, τα αρχεία συμπιέζονται κάθε φορά που τα αποθηκεύετε, διατηρώντας έτσι τα αρχεία σε πολύ μικρότερο μέγεθος από ό,τι οι μορφές εγγράφων των προηγούμενων εκδόσεων του Microsoft Office.

Η μορφή των Open Office XML αρχείων είναι αρθρωτή. Έτσι διαφορετικοί τύποι πληροφοριών αποθηκεύονται ανεξάρτητα. Κάθε έγγραφο είναι στην πραγματικότητα ένα πακέτο ZIP που περιέχει αρκετά αρχεία, ένα για το κυρίως σώμα του εγγράφου, ένα για τις κεφαλίδες και ένα για τα σχόλια. Έτσι, αν παρουσιαστεί κάποιο σφάλμα σε ένα στοιχείο του εγγράφου, τα υπόλοιπα στοιχεία δεν επηρεάζονται, γεγονός που απλοποιεί πολύ την ανάκτηση του περιεχομένου.

Η αλλαγή από το format (.doc) στο νέο (.docx) είχε ως επακόλουθο τα προγράμματα που ανέλυαν metadata στις προηγούμενες εκδόσεις του MS-Office, να μην μπορούν να αναλύσουν metadata βασισμένα στο OOXML πρότυπο.

Η παρούσα διπλωματική ασχολείται με τα μεταδεδομένα που είναι διαθέσιμα στο νέο πρότυπο.

Ένα αρχείο Office Open XML, αποτελείται από μια σειρά σχετιζόμενων μεταξύ τους στοιχείων, τα οποία αποθηκεύονται σε πακέτα ZIP.

Πληροφορίες για τις σχέσεις των στοιχείων με τα πακέτα, αποθηκεύονται ως αντικείμενο μέσα στο πακέτο.

4.1 Διάρθρωση εγγράφου OOXML

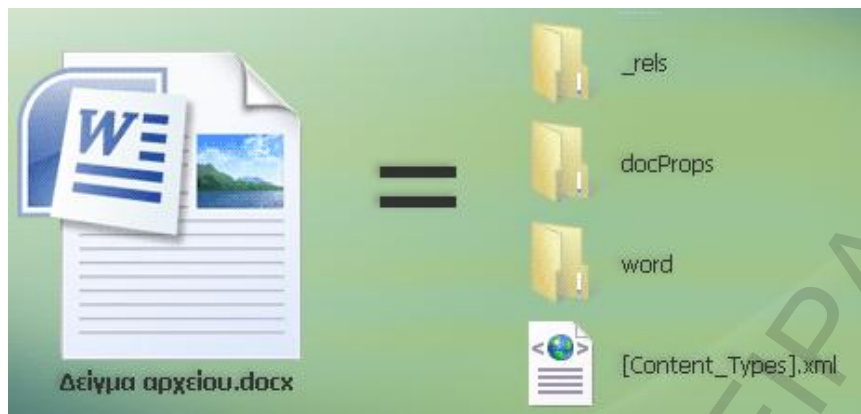
Ο φάκελος **_rels** περιέχει ένα αρχείο που ονομάζεται **.rels** και το οποίο αποθηκεύει πληροφορίες σχετικά με τις σχέσεις μεταξύ των στοιχείων που υπάρχουν στο πακέτο ZIP. Μέσω του αρχείου **.rels** τα προγράμματα του Office system γνωρίζουν πού θα βρουν τα στοιχεία του εγγράφου όταν ανοίγουν ένα έγγραφο.

Ο κύριος φάκελος του εγγράφου, (φάκελος **word**), αποθηκεύει το κύριο περιεχόμενο του εγγράφου, τυχόν πολυμέσα (όπως φωτογραφίες) που περιέχονται μέσα στο έγγραφο, καθώς και διάφορα στοιχεία του εγγράφου, όπως ρυθμίσεις, κεφαλίδες και θέματα. Ο κύριος φάκελος του εγγράφου περιέχει επίσης τον δικό του φάκελο **_rels**, όπου ορίζονται οι σχέσεις μεταξύ των στοιχείων που βρίσκονται μέσα στον κύριο φάκελο του εγγράφου.

⁵ <http://www.ecma-international.org/memento/TC45.htm>

⁶ [http://msdn.microsoft.com/en-us/library/gg607163\(v=office.14\).aspx](http://msdn.microsoft.com/en-us/library/gg607163(v=office.14).aspx)

Το αρχείο **[Content_Types].xml** περιέχει ορισμούς των τύπων περιεχομένου που περιέχονται στο πακέτο ZIP, όπως το κύριο έγγραφο, το θέμα του εγγράφου και τις ιδιότητες του αρχείου. Αυτό το αρχείο αποθηκεύει επίσης ορισμούς των επεκτάσεων των αρχείων που χρησιμοποιούνται στο πακέτο ZIP, όπως τις μορφές αρχείων (για παράδειγμα, png ή jpeg) των εικόνων που περιλαμβάνονται σε ένα έγγραφο.



4.2 Πλεονεκτήματα των μορφών Open XML

Η μορφή Open XML διαθέτει πλήθος πλεονεκτημάτων⁷ — όχι μόνο για τους προγραμματιστές και τις λύσεις που αναπτύσσουν, αλλά και για μεμονωμένα άτομα ή εταιρείες κάθε μεγέθους:

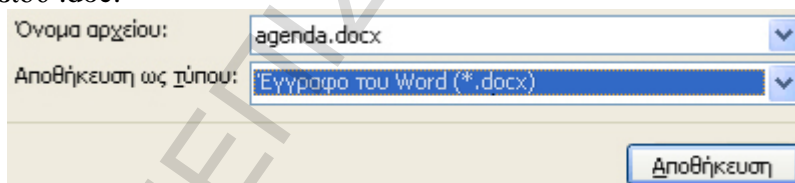
- **Συμπίεση αρχείων** Τα αρχεία συμπιέζονται αυτόματα και σε ορισμένες περιπτώσεις μπορούν να είναι κατά 75% μικρότερα. Η μορφή Open XML χρησιμοποιεί την τεχνολογία συμπίεσης zip για την αποθήκευση εγγράφων, προσφέροντας πιθανή μείωση του κόστους, καθώς μειώνεται ο χώρος στο δίσκο που απαιτείται για την αποθήκευση των αρχείων και ελαχιστοποιείται το εύρος ζώνης που απαιτείται για την αποστολή αρχείων μέσω ηλεκτρονικού ταχυδρομείου σε δίκτυα ή στο Internet. Όταν ανοίγετε ένα αρχείο, το αρχείο αποσυμπίεζεται αυτόματα. Όταν αποθηκεύετε ένα αρχείο, το αρχείο συμπιέζεται και πάλι αυτόματα. Δεν απαιτείται η εγκατάσταση ειδικών προγραμμάτων συμπίεσης για το άνοιγμα και το κλείσιμο των αρχείων του Office.
- **Βελτιωμένη ανάκτηση κατεστραμμένων αρχείων** Τα αρχεία διαθέτουν αρθρωτή δομή που διατηρεί διαφορετικά στοιχεία δεδομένων στο αρχείο ξεχωριστά. Αυτό επιτρέπει το άνοιγμα ενός αρχείου έστω και εάν ένα στοιχείο μέσα στο αρχείο (π.χ. ένα γράφημα ή ένας πίνακας) έχει καταστραφεί ή αλλοιωθεί.
- **Καλύτερη προστασία προσωπικών δεδομένων και περισσότερος έλεγχος σχετικά με προσωπικές πληροφορίες** Μπορείτε να μοιράζεστε έγγραφα με ασφάλεια, καθώς τα προσωπικά και επιχειρηματικά δεδομένα, όπως το όνομα του συντάκτη, τα σχόλια, η παρακολούθηση αλλαγών και οι διαδρομές αρχείων αναγνωρίζονται εύκολα και αφαιρούνται με τη χρήση του ελέγχου εγγράφου.

⁷ <http://office.microsoft.com/el-gr/powerpoint-help/HA010354237.aspx>

- Καλύτερη ενοποίηση και διαλειτουργικότητα επιχειρηματικών δεδομένων Η χρήση της μορφής Open XML ως το πλαίσιο διαλειτουργικότητας για το σύνολο των προϊόντων του Office σημαίνει ότι τα έγγραφα, τα φύλλα εργασίας, οι παρουσιάσεις και οι φόρμες μπορούν να αποθηκευθούν σε μορφή αρχείου XML που μπορεί να χρησιμοποιήσει ελεύθερα ο καθένας, χωρίς προβλήματα αδείας. Το Office υποστηρίζει επίσης προσαρμοσμένες διατάξεις XML που βελτιώνουν τους υπάρχοντες τύπους εγγράφων του Office. Αυτό σημαίνει ότι οι πελάτες μπορούν εύκολα να δουν πληροφορίες στα υπάρχοντα συστήματα και να τις επεξεργαστούν σε προγράμματα του Office με τα οποία είναι εξοικειωμένοι. Οι πληροφορίες που δημιουργούνται μέσα στο Office μπορούν εύκολα να χρησιμοποιηθούν από άλλες επιχειρηματικές εφαρμογές. Το μόνο που χρειάζεται για να ανοίξετε και να επεξεργαστείτε ένα αρχείο του Office είναι ένα βοηθητικό πρόγραμμα συμπίεσης ZIP και ένα πρόγραμμα επεξεργασίας XML.
- Ευκολότερος εντοπισμός ενός εγγράφου που περιέχει μακροεντολές Τα αρχεία που αποθηκεύονται με το προεπιλεγμένο επίθημα "x" (όπως .docx, .xlsx και .pptx) δεν μπορούν να περιέχουν μακροεντολές Visual Basic for Applications (VBA) και μακροεντολές XML. Μόνο τα αρχεία των οποίων η επέκταση ονόματος αρχείου λήγει σε "m" (όπως .docm, .xlsm και .pptm) μπορούν να περιέχουν μακροεντολές.

4.3 Επεκτάσεις ονόματος αρχείου XML

Από προεπιλογή, τα έγγραφα, τα φύλλα εργασίας και οι παρουσιάσεις που δημιουργείτε στο Office αποθηκεύονται σε μορφή XML με επεκτάσεις ονόματος αρχείου που προσθέτουν ένα "x" ή ένα "m" στις επεκτάσεις ονόματος αρχείου που γνωρίζετε ήδη. Το "x" υποδηλώνει ένα αρχείο XML που δεν διαθέτει μακροεντολές και το "m" υποδηλώνει ένα αρχείο XML που περιέχει μακροεντολές. Για παράδειγμα, όταν αποθηκεύετε ένα έγγραφο στο Word, το αρχείο χρησιμοποιεί πλέον την επέκταση ονόματος αρχείου .docx από προεπιλογή, αντί της επέκτασης ονόματος αρχείου .doc.



Όταν αποθηκεύετε ένα αρχείο ως πρότυπο, βλέπετε το ίδιο είδος αλλαγής. Η επέκταση προτύπου που χρησιμοποιούνταν σε προηγούμενες εκδόσεις υπάρχει και εδώ, αλλά τώρα διαθέτει ένα "x" ή ένα "m" στο τέλος. Εάν το αρχείο περιέχει κώδικα ή μακροεντολές, θα πρέπει να το αποθηκεύσετε χρησιμοποιώντας τη νέα μορφή αρχείου XML με δυνατότητα μακροεντολών, η οποία προσθέτει στην επέκταση του αρχείου ένα "m" για τις μακροεντολές.

Στους παρακάτω πίνακες παρατίθενται όλες οι προεπιλεγμένες επεκτάσεις ονόματος αρχείου στο Word, το Excel και το PowerPoint.

4.3.1 Word

ΤΥΠΟΣ ΑΡΧΕΙΟΥ XML	ΕΠΕΚΤΑΣΗ
Έγγραφο	.docx

Έγγραφο με μακροεντολές	.docm
Πρότυπο	.dotx
Πρότυπο με μακροεντολές	.dotm

4.3.2 Excel

ΤΥΠΟΣ ΑΡΧΕΙΟΥ XML	ΕΠΕΚΤΑΣΗ
Βιβλίο εργασίας	.xlsx
Βιβλίο εργασίας με μακροεντολές	.xlsm
Πρότυπο	.xltx
Πρότυπο με μακροεντολές	.xltm
Διαδίκτυο βιβλίο εργασίας που δεν είναι του τύπου XML	.xlsb
Πρόσθετο με μακροεντολές	.xlam

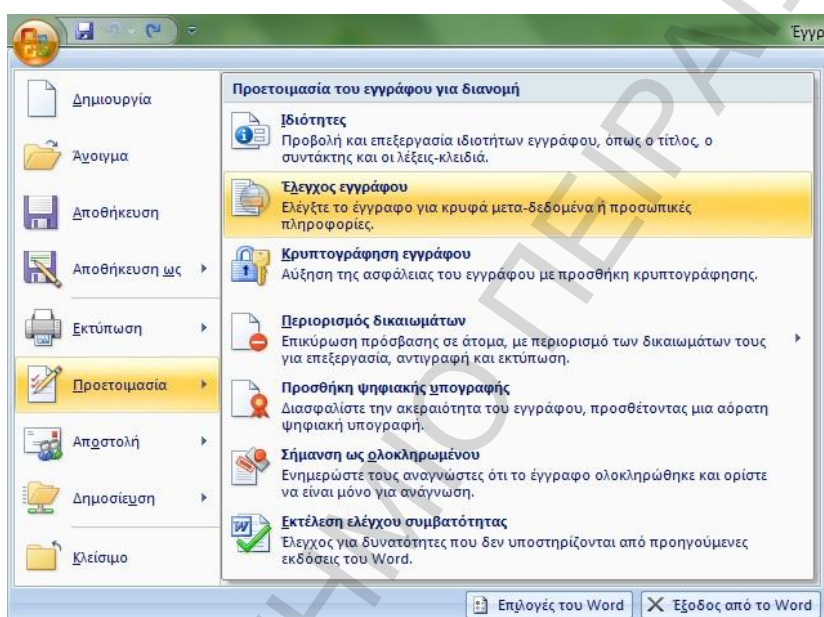
4.3.3 PowerPoint

ΤΥΠΟΣ ΑΡΧΕΙΟΥ XML	ΕΠΕΚΤΑΣΗ
Παρουσίαση	.pptx
Παρουσίαση με μακροεντολές	.pptm
Πρότυπο	.potx
Πρότυπο με μακροεντολές	.potm
Πρόσθετο με μακροεντολές	.ppam
Εμφάνιση	.ppsx
Εμφάνιση με μακροεντολές	.ppsm
Διαφάνεια	.sldx
Διαφάνεια με μακροεντολές	.sldm
Θέμα του Office	.thmx

5 Εργαλεία Εύρεσης, Ανάλυσης ή Αφαίρεσης Μεταδεδομένων

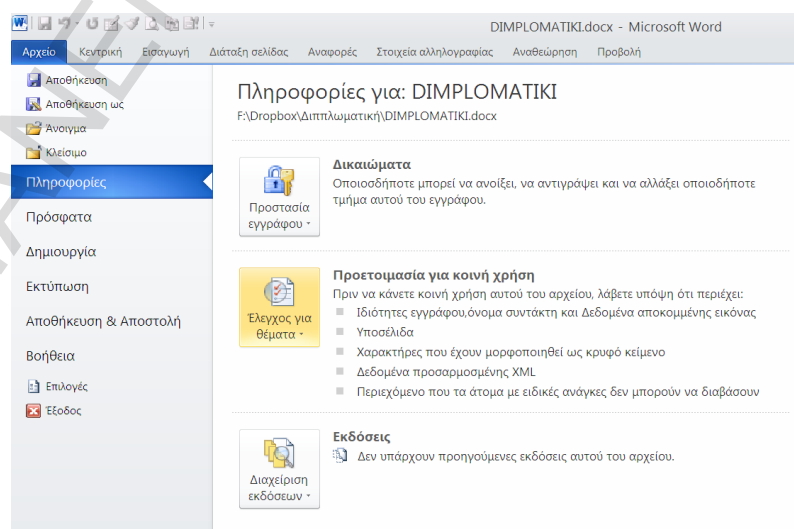
5.1 Κατάργηση κρυφών δεδομένων και προσωπικών πληροφοριών με επιθεώρηση εγγράφων

Ο Έλεγχος εγγράφου εμφανίζει διαφορετικά σύνολα επιθεωρήσεων στο Office Word 2007, Office Excel 2007, και στο Office PowerPoint 2007 δίνοντας τη δυνατότητα εύρεσης και κατάργησης κρυφών δεδομένων και προσωπικών πληροφοριών που αφορούν συγκεκριμένα κάθε ένα από αυτά τα προγράμματα.



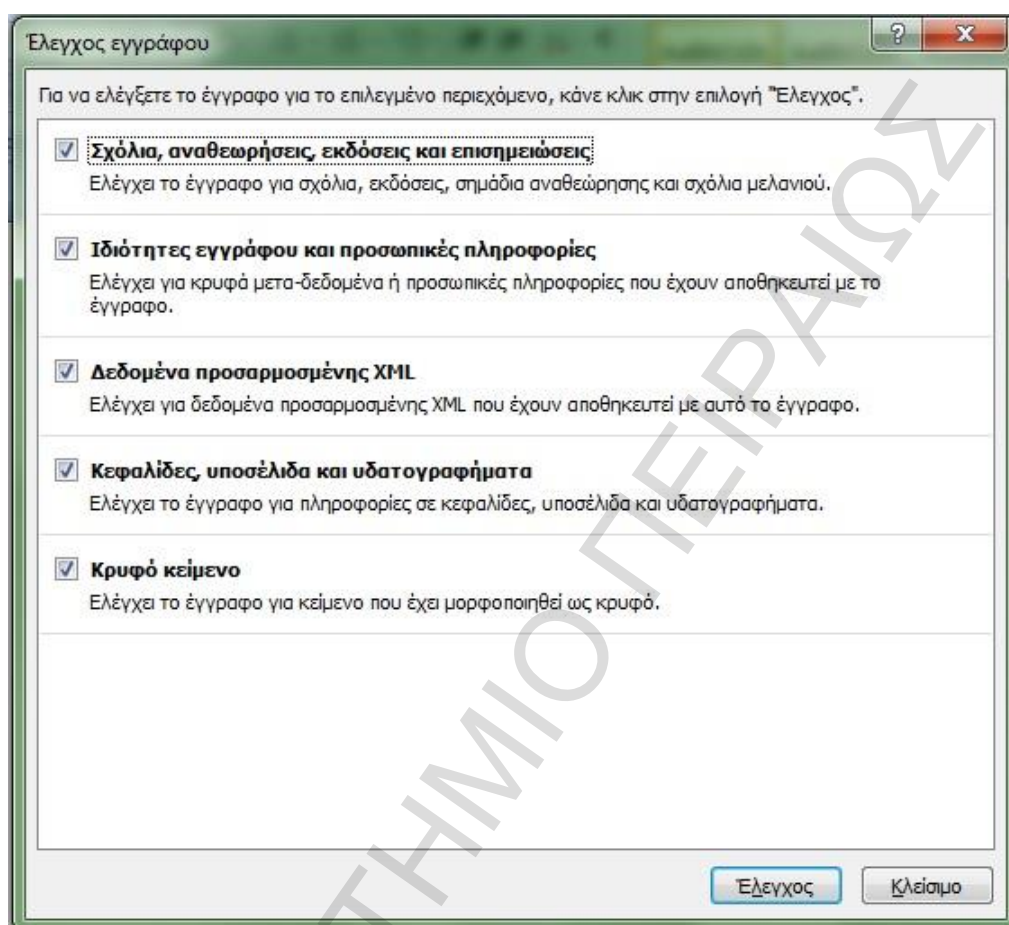
Εικόνα 1 Έναρξη του στοιχείου "Έλεγχος Εγγράφου" στο Office 2007

Ο Έλεγχος εγγράφου βρίσκει και καταργεί κρυφά δεδομένα και προσωπικές πληροφορίες σε έγγραφα του Office που έχουν δημιουργηθεί στα Office Word 2007, Office Excel 2007, Office PowerPoint 2007, και σε προηγούμενες εκδόσεις αυτών των προγραμμάτων.



Εικόνα 2 Έναρξη του στοιχείου "Έλεγχος Εγγράφου" στο Office 2010

Στο παράθυρο διαλόγου επιθεώρηση εγγράφου, μπορούν να ενεργοποιηθούν τα πλαίσια ελέγχου και να επιλεγθούν οι τύποι κρυφού περιεχομένου προς έλεγχο και στη συνέχεια προς κατάργηση.



Εικόνα 3 Πλαίσιο διαλόγου του Ελέγχου Εγγράφου στο MS Word 2007

Μπορεί κανείς να απενεργοποιήσει τη λειτουργία Ελέγχου Εγγράφου, πραγματοποιώντας μια τοπική ρύθμιση στον προσωπικό υπολογιστή του χρήστη.

Κατάργηση κρυφών δεδομένων και προσωπικών πληροφοριών με επιθεώρηση εγγράφων:

Ανά πεδίο ο έλεγχος που γίνεται αφορά⁸:

ΟΝΟΜΑ ΕΛΕΓΧΟΥ	ΒΡΙΣΚΕΙ ΚΑΙ ΚΑΤΑΡΓΕΙ
Σχόλια, αναθεωρήσεις και εκδόσεις	<ul style="list-style-type: none"> • Σχόλια • Σημάδια αναθεώρησης από παρακολούθηση αλλαγών • Πληροφορίες έκδοσης εγγράφου • Σχόλια μελανιού
Ιδιότητες εγγράφου και προσωπικές	<ul style="list-style-type: none"> • Ιδιότητες εγγράφου, συμπεριλαμβανομένων πληροφοριών από τις καρτέλες Σύνοψη, Στατιστικά

⁸ <http://office.microsoft.com/el-gr/word-help/HA010354329.aspx>

πληροφορίες

στοιχεία και Προσαρμογή του παραθύρου διαλόγου Ιδιότητες εγγράφου

- Κεφαλίδες ηλεκτρονικού ταχυδρομείου
- Δελτία δρομολόγησης
- Πληροφορίες αποστολής για αναθεώρηση
- Ιδιότητες διακομιστή εγγράφων
- Πληροφορίες πολιτικής διαχείρισης εγγράφων
- Πληροφορίες τύπου περιεχομένου
- Πληροφορίες σύνδεσης δεδομένων για πεδία που βασίζονται σε δεδομένα (η τελευταία τιμή μετατρέπεται σε κείμενο)
- Όνομα χρήστη
- Όνομα προτύπου

Κεφαλίδες, υποσέλιδα και υδατογραφήματα

- Πληροφορίες σε κεφαλίδες εγγράφων
- Πληροφορίες σε υποσέλιδα εγγράφων
- Υδατογραφήματα

Κρυφό κείμενο

- Κείμενο μορφοποιημένο ως κρυφό (εφέ γραμματοσειράς που είναι διαθέσιμο στο παράθυρο διαλόγου Γραμματοσειρά)

Δεν μπορεί να εντοπίσει κείμενο στο οποίο έχει εφαρμοστεί απόκρυψη με άλλες μεθόδους (για παράδειγμα, λευκό κείμενο σε λευκό φόντο).

Δεδομένα προσαρμοσμένης XML

- Δεδομένα προσαρμοσμένης XML που μπορούν να αποθηκευτούν σε ένα έγγραφο

Αόρατο περιεχόμενο

- Αντικείμενα που δεν φαίνονται, επειδή έχουν μορφοποιηθεί ως μη ορατά.

Δεν μπορεί να εντοπίσει αντικείμενα που καλύπτονται από άλλα αντικείμενα

5.2 Κατάργηση κρυφών δεδομένων και προσωπικών πληροφοριών με έλεγχο βιβλίων εργασίας⁹

ΟΝΟΜΑ ΕΛΕΓΧΟΥ	ΒΡΙΣΚΕΙ ΚΑΙ ΚΑΤΑΡΓΕΙ
Σχόλια και επισημειώσεις	<ul style="list-style-type: none">• Σχόλια• Σχόλια με γραφή
Ιδιότητες εγγράφου	<ul style="list-style-type: none">• Ιδιότητες εγγράφου, συμπεριλαμβανομένων

⁹ <http://office.microsoft.com/el-gr/excel-help/HA010354331.aspx?CTT=1>

και Προσωπικές πληροφορίες

πληροφοριών από τις καρτέλες **Σύνοψη, Στατιστικά στοιχεία** και **Προσαρμογή** του παραθύρου διαλόγου **Ιδιότητες εγγράφου**

- Θέση εγγράφου,
- Κεφαλίδες ηλεκτρονικού ταχυδρομείου
- Δελτία δρομολόγησης
- Πληροφορίες αποστολής για αναθεώρηση
- Ιδιότητες διακομιστή εγγράφων
- Πληροφορίες πολιτικής διαχείρισης εγγράφων
- Πληροφορίες τύπου περιεχομένου
- Όνομα χρήστη
- Πληροφορίες διαδρομής εκτυπωτή
- Σχόλια σεναρίου
- Διαδρομή αρχείου για δημοσίευση ιστοσελίδων
- Σχόλια για καθορισμένα ονόματα και ονόματα πινάκων
- Ανενεργές συνδέσεις με εξωτερικά δεδομένα

Κεφαλίδες και υποσέλιδα

- Πληροφορίες σε κεφαλίδες φύλλων εργασίας
- Πληροφορίες σε υποσέλιδα φύλλων εργασίας

Κρυφές γραμμές και στήλες

- Κρυφές γραμμές
- Κρυφές στήλες που περιέχουν δεδομένα
- Αν στο βιβλίο εργασίας υπάρχουν κρυφές στήλες που δεν περιέχουν δεδομένα και βρίσκονται μεταξύ στηλών οι οποίες περιέχουν δεδομένα, οι κενές αυτές κρυφές στήλες θα εντοπίζονται και θα καταργούνται.
- Αν οι κρυφές γραμμές ή στήλες στο βιβλίο εργασίας σας περιέχουν δεδομένα, μπορείτε να αλλάξετε τα αποτελέσματα των υπολογισμών ή των τύπων στο βιβλίο εργασίας καταργώντας τις. Αν δεν γνωρίζετε τις πληροφορίες που περιέχονται στις κρυφές γραμμές ή στήλες, κλείστε τον Έλεγχο εγγράφου, καταργήστε την απόκρυψη των κρυφών γραμμών ή στηλών και στη συνέχεια αναθεωρήστε τα περιεχόμενά τους.
- Αυτός ο Έλεγχος δεν εντοπίζει σχήματα, γραφήματα, στοιχεία ελέγχου, αντικείμενα και στοιχεία ελέγχου Microsoft ActiveX, εικόνες ή γραφικά SmartArt τα οποία ενδέχεται να βρίσκονται σε κρυφές στήλες.

Κρυφά φύλλα εργασίας

- Κρυφά φύλλα εργασίας

Αν τα κρυφά φύλλα εργασίας στο βιβλίο εργασίας σας

περιέχουν δεδομένα, μπορείτε να αλλάξετε τα αποτελέσματα των υπολογισμών ή των τύπων στο βιβλίο εργασίας σας καταργώντας τα. Αν δεν γνωρίζετε ποιες πληροφορίες περιέχονται στα κρυφά φύλλα εργασίας, κλείστε τον Έλεγχο εγγράφων, καταργήστε την απόκρυψη των κρυφών φύλλων εργασίας και κατόπιν αναθεωρήστε τα περιεχόμενά τους.

Προσαρμοσμένα δεδομένα XML

- Προσαρμοσμένα δεδομένα XML που μπορεί να είναι αποθηκευμένα σε ένα βιβλίο εργασίας

Αόρατο περιεχόμενο

- Αντικείμενα που δεν φαίνονται επειδή έχουν μορφοποιηθεί ως αόρατα

Ο έλεγχος αυτός δεν εντοπίζει αντικείμενα που καλύπτονται από άλλα αντικείμενα.

5.3 Κατάργηση κρυφών δεδομένων και προσωπικών πληροφοριών με έλεγχο παρουσιάσεων¹⁰

ΟΝΟΜΑ ΕΠΙΘΕΩΡΗΣΗΣ	ΒΡΙΣΚΕΙ ΚΑΙ ΚΑΤΑΡΓΕΙ
Σχόλια και επισημειώσεις	<ul style="list-style-type: none"> • Σχόλια • Σχόλια με μελάνι
Ιδιότητες εγγράφου και προσωπικές πληροφορίες	<ul style="list-style-type: none"> • Ιδιότητες εγγράφου, συμπεριλαμβανομένων πληροφοριών από τις καρτέλες Σύνοψη, Στατιστικά στοιχεία και Προσαρμογή του παραθύρου διαλόγου Ιδιότητες εγγράφου • Κεφαλίδες ηλεκτρονικού ταχυδρομείου • Δελτία δρομολόγησης • Πληροφορίες αποστολής για αναθεώρηση • Ιδιότητες διακομιστή εγγράφων • Πληροφορίες πολιτικής διαχείρισης εγγράφων • Πληροφορίες τύπου περιεχομένου • Διαδρομή αρχείου για δημοσίευση ιστοσελίδων
Αόρατο περιεχόμενο εντός διαφάνειας	<ul style="list-style-type: none"> • Αντικείμενα που δεν φαίνονται επειδή έχουν μορφοποιηθεί ως μη-ορατά

Η Επιθεώρηση δεν εντοπίζει αντικείμενα που

¹⁰ <http://office.microsoft.com/el-gr/powerpoint-help/HA010354330.aspx?CTT=1>

καλύπτονται από άλλα αντικείμενα.

Περιεχόμενο εκτός διαφάνειας

- Περιεχόμενο ή αντικείμενα που δεν φαίνονται αμέσως σε μια παρουσίαση επειδή βρίσκονται έξω από την περιοχή της διαφάνειας, συμπεριλαμβανομένων των εξής:
- Εικόνες clip art
- Πλαίσια κειμένου
- Γραφικά
- Πίνακες

Η δυνατότητα Επιθεώρηση εγγράφου δεν εντοπίζει ή καταργεί αντικείμενα με εφέ κίνησης που βρίσκονται εκτός διαφάνειας.

Σημειώσεις παρουσίασης

- Κείμενο που προστέθηκε στην ενότητα "Σημειώσεις" μιας παρουσίασης

Η δυνατότητα "Επιθεώρηση εγγράφου" δεν είναι δυνατόν να καταργήσει εικόνες που προστίθενται στην ενότητα "Σημειώσεις" μιας παρουσίασης.


Δεδομένα προσαρμοσμένης XML

- Δεδομένα προσαρμοσμένης XML που μπορούν να αποθηκευτούν σε μια παρουσίαση

5.4 Επισκόπηση βασικών Ιδιοτήτων Ελέγχου Εγγράφου

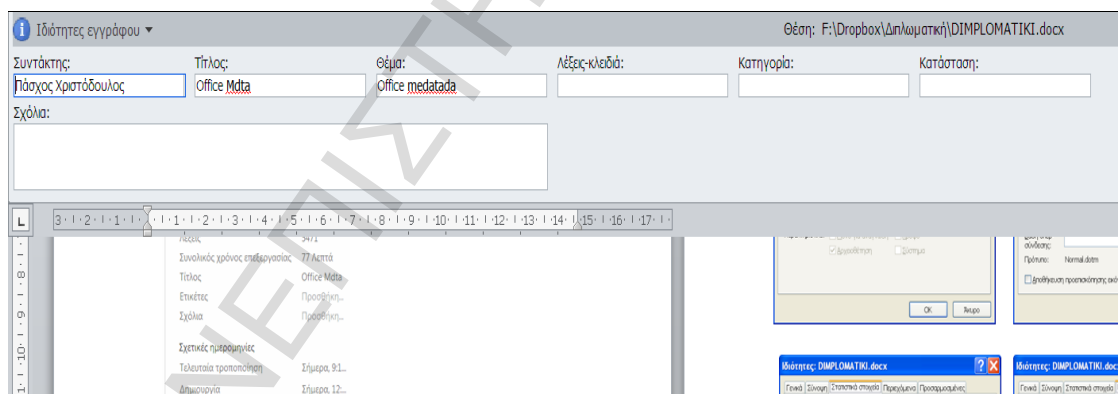
Οι βασικές ιδιότητες Ελέγχου Εγγράφου μας δίνουν πληροφορίες για το έγγραφο που έχουμε ανοίξει. Από αυτές ο χρήστης μπορεί να τροποποιήσει τις:

- Τίτλος
- Ετικέτες
- Σχόλια
- Συντάκτης (μπορεί να προστεθεί Συντάκτης)

Ιδιότητες ▾	
Μέγεθος	1,99MB
Σελίδες	43
Λέξεις	5471
Συνολικός χρόνος επεξεργασίας	77 Λεπτά
Τίτλος	Office Mdata
Ετικέτες	Προσθήκη...
Σχόλια	Προσθήκη...
Σχετικές ημερομηνίες	
Τελευταία τροποποίηση	Σήμερα, 9:1...
Δημιουργία	Σήμερα, 12:...
Τελευταία εκτύπωση	Ποτέ
Σχετικά άτομα	
Συντάκτης	Πάσχος Χρ...
	Προσθήκη...
Τελευταία τροποποίηση από	Chris
Σχετικά έγγραφα	
 Ανοιγμα θέσης αρχείου	
Εμφάνιση όλων των ιδιοτήτων	

Εικόνα 4 Προβολή των Βασικών Ιδιοτήτων Ελέγχου Εγγράφου

Οι βασικές ιδιότητες μπορεί να είναι διαθέσιμες και μέσα στο κυρίως έγγραφο με τη μορφή γραμμής εργαλείων:



Εικόνα 5 Προβολή Βασικών Ιδιοτήτων Ελέγχου Εγγράφου

5.5 Τι δεν ανιχνεύει ο Έλεγχος εγγράφου;

5.5.1 Κρυφό ή Αόρατο Περιεχόμενο

Το Excel, το PowerPoint και το Word έχουν λειτουργίες του Ελέγχου Εγγράφου που αφαιρούν ορισμένους τύπους κρυφών ή αόρατων δεδομένων. Το Excel έχει τρεις λειτουργίες για την αφαίρεση αόρατων αντικειμένων, κρυφών σηλών, κρυφών

γραμμών και κρυφών φύλλων εργασίας. Το PowerPoint έχει μια ακόμη λειτουργία για την αφαίρεση αόρατου περιεχομένου στην διαφάνεια, ενώ το Word έχει μια λειτουργία για την αφαίρεση κρυφού κειμένου.

Οι λειτουργίες του Ελέγχου Εγγράφου για κρυφό και αόρατο περιεχόμενο δεν απομακρύνουν κανένα περιεχόμενο, εκτός αν αυτό έχει μορφοποιηθεί διακριτά ως κρυφό ή αόρατο.

Για παράδειγμα, εάν τοποθετηθεί άσπρο κείμενο σε άσπρο φόντο, το κείμενο κρύβεται αποτελεσματικά, όμως ο Έλεγχος Εγγράφου θεωρεί ότι αυτό έγινε ηθελημένα και δεν το αφαιρεί. Το ίδιο συμβαίνει εάν δημιουργηθεί ένα μπλε σχήμα σε μπλε φόντο ή αν ένα σχήμα έχει καλυφθεί από ένα άλλο ή αν ένα σχήμα δεν έχει γέμισμα ούτε περίγραμμα, παρόλο που δεν είναι ορατά.

Επίσης, μπορεί κανείς να κρύψει δεδομένα στο Excel τοποθετώντας τα σε μια μακρινή στήλη ή γραμμή, όπως η γραμμή 10.000 ή η στήλη 1.000.000. Έτσι κρύβονται αποτελεσματικά τα δεδομένα καθώς είναι εκτός οθόνης ή πολύ μακριά από την τυπική περιοχή προβολής ενός φύλλου εργασίας. Όμως ο Έλεγχος Εγγράφου τα βλέπει και τα θεωρεί κανονικά δεδομένα και δεν τα αφαιρεί. Το ίδιο ισχύει και για σχήματα ή SmartArt που μετακινείται πέρα από την ορατή περιοχή προβολής.

Αξίζει να σημειωθεί ότι το PowerPoint, αντίθετα με το Excel και το Word, έχει μια λειτουργία Ελεγκτή Εγγράφου που αφαιρεί περιεχόμενο που βρίσκεται εκτός οθόνης, ακόμη κι αν αυτό έχει σημειωθεί ως κρυφό ή αόρατο.

5.5.2 Δεδομένα Προσωρινής Αποθήκευσης

Αρκετά χαρακτηριστικά ενός προγράμματος βασίζονται στην προσωρινή αποθήκευση δεδομένων προκειμένου να αυξήσουν την επίδοση. Αυτό μπορεί να είναι πρόβλημα εάν αυτά τα δεδομένα περιέχουν μεταδεδομένα καθώς ο Έλεγχος εγγράφου δεν τα αφαιρεί από τα αρχεία.

Οι συγκεντρωτικοί πίνακες είναι ένα παράδειγμα όπου ο Έλεγχος Εγγράφου δεν απομακρύνει προσωρινώς αποθηκευμένα δεδομένα. Όταν δημιουργεί κανείς έναν Συγκεντρωτικό Πίνακα σε ένα νέο φύλλο εργασίας, το Excel αποθηκεύει προσωρινά τα δεδομένα που επιλέγονται. Σε κάποιες περιπτώσεις αυτά τα δεδομένα μπορεί να παραμείνουν στο φύλλο εργασίας ακόμη κι αν ο πίνακας διαγραφεί. Προκειμένου να αφαιρεθούν πρέπει να από-επιλεχθεί το πλαίσιο "Αποθήκευση δεδομένων προέλευσης με το αρχείο" που βρίσκεται στην καρτέλα "Δεδομένα" στις "Επιλογές Συγκεντρωτικού Πίνακα".

Η χρήση των λειτουργιών ταξινόμησης και φίλτρου μπορεί επίσης να δημιουργήσει προσωρινώς αποθηκευμένα δεδομένα. Γενικά, αυτό δεν είναι πρόβλημα καθώς τα δεδομένα αυτά προέρχονται από ορατά δεδομένα του φύλλου εργασίας, αλλά είναι πιθανό να μην υπάρχουν πλέον στο φύλλο εργασίας. Προφανώς, ο Έλεγχος Εγγράφου δεν αφαιρεί τέτοιου είδους δεδομένα προσωρινής αποθήκευσης από ένα φύλλο εργασίας.

Τα ενσωματωμένα αντικείμενα (embedded objects) μπορούν να είναι επίσης πηγή δεδομένων προσωρινής αποθήκευσης. Για παράδειγμα, εάν αντιγραφεί ένα διάγραμμα από το Excel και χρησιμοποιηθεί η προεπιλεγμένη επιλογή επικόλλησης για να επικολληθεί σε μια διαφάνεια του PowerPoint, στην πραγματικότητα επικολλούνται και τα υποκείμενα δεδομένα από τα οποία προέρχεται το διάγραμμα. Αν και το διάγραμμα είναι ορατό, τα σχετιζόμενα δεδομένα δεν είναι ορατά, αλλά είναι προσωρινώς αποθηκευμένα. Η αφαίρεση ή διαγραφή του διαγράμματος δεν αφαιρεί απαραίτητα τα δεδομένα αυτά και ο Έλεγχος Εγγράφου δεν τα απομακρύνει. Γενικά, ο Έλεγχος δεν αφαιρεί δεδομένα που σχετίζονται με ενσωματωμένα αντικείμενα. Προκειμένου κατά την επικόλληση να μην συμπεριληφθούν σχετικά δεδομένα με το αντικείμενο, πρέπει να χρησιμοποιηθεί η επιλογή Επικόλληση ως Εικόνα.

5.5.3 Εξωτερικές συνδέσεις

Οι συνδέσεις βάσης δεδομένων και εκτυπωτή είναι δύο συνήθεις τύποι εξωτερικών συνδέσεων που μπορούν να τοποθετήσουν μεταδεδομένα σε ένα αρχείο χωρίς να το γνωρίζει κανείς. Και στις δύο περιπτώσεις ο Έλεγχος Εγγράφου δεν μπορεί να αφαιρέσει αυτήν την πληροφορία από το αρχείο.

Οι συνδέσεις δεδομένων μπορεί να είναι ιδιαίτερα περίπλοκες, επειδή συνήθως πρέπει να παρέχεται ιδιωτική πληροφορία προκειμένου να δημιουργηθεί η σύνδεση, όπως το όνομα χρήστη, ο κωδικός πρόσβασης, η διαδρομή ως τη βάση, το όνομα της βάσης και το όνομα του μηχανήματος από το οποίο δημιουργείται η σύνδεση. Αυτή η ιδιωτική πληροφορία δημιουργεί τη συμβολοσειρά της σύνδεσης η οποία αποθηκεύεται στο αρχείο Excel. Παρόλα αυτά διαγράφεται εάν διαγραφεί η σύνδεση. Επίσης, οι ιδιότητες σύνδεσης μπορούν να διαμορφωθούν ώστε οι κωδικοί πρόσβασης να μην αποθηκεύονται με την πληροφορία σύνδεσης, κάτι που συστήνεται ως καλή πρακτική.

Η πληροφορία εκτυπωτή είναι επίσης περίπλοκη επειδή οι εφαρμογές του Office περνούν πληροφορία σχετική με τον εκτυπωτή στους οδηγούς εκτυπωτών και το κάνουν αυτό ενσωματώνοντας την πληροφορία στο έγγραφο, φύλλο εργασίας ή αρχείο παρουσίασης. Αυτή η πληροφορία μπορεί να περιλαμβάνει τη διαδρομή ως τον εκτυπωτή και το όνομα εκτυπωτή. Μπορεί επίσης να περιλαμβάνει όνομα χρήστη και κωδικό πρόσβασης εάν χρησιμοποιούνται λειτουργίες ασφαλούς εκτύπωσης. Ο Έλεγχος Εγγράφου μπορεί να αφαιρέσει το όνομα του εκτυπωτή και την πληροφορία διαδρομής από ένα αρχείο, αλλά δεν μπορεί να αφαιρέσει όλη την πληροφορία σχετικά με το συγκεκριμένο εκτυπωτή, καθώς οι οδηγοί εκτυπωτή συνήθως δεν παρέχουν αρκετά λεπτομερή πληροφορία ώστε να αποφασίσει ο Έλεγχος τι τύπου μεταδεδομένα είναι ενσωματωμένα σε ένα αρχείο.

5.5.4 Προστατευμένα και περιορισμένα αρχεία

Ο Έλεγχος Εγγράφου δεν αφαιρεί μεταδεδομένα που βρίσκονται σε ένα προστατευμένο ή περιορισμένο αρχείο, όπως ένα αρχείο που έχει περιορισμούς επεξεργασίας, είναι ψηφιακά υπογεγραμμένο ή προστατεύεται από περιορισμένες

άδειες, αφού δεν μπορεί να έχει πρόσβαση στο αρχείο. Κατά κανόνα πρέπει να τρέξει κανείς τον Έλεγχο πριν περιορίσει ή προστατεύσει ένα αρχείο.

Επιπροσθέτως, ο Έλεγχος Εγγράφου δεν αφαιρεί σχόλια που προστίθενται από ένα χρήστη όταν εφαρμόζει μια ψηφιακή υπογραφή σε ένα έγγραφο. Από τη στιγμή που δεν υπάρχει έλεγχος στο τι μπορεί να πει σε σχόλιο κάποιος που υπογράφει ψηφιακά, μπορεί να περιέχονται εκεί μεταδεδομένα που να μην είναι επιθυμητή η αποκάλυψή τους. Αυτό συμβαίνει όταν εισάγεται μια γραμμή ψηφιακής υπογραφής του Microsoft Office και τσεκάρεται το πλαίσιο "Να επιτρέπεται η προσθήκη σχολίων από τον υπογράφοντα" στο παράθυρο διαλόγου "Υπογραφή". Αυτή η επιλογή επιτρέπει έναν υπογράφοντα να δημιουργήσει ένα σχόλιο όταν προσθέτει την υπογραφή του σε ένα έγγραφο. Οποιοσδήποτε μπορεί να δει το σχόλιο κοιτώντας τις λεπτομέρειες της υπογραφής. Επειδή όμως το έγγραφο έχει υπογραφεί ψηφιακά και δεν μπορεί να τροποποιηθεί, δεν μπορεί να αφαιρεθεί το σχόλιο από τον Έλεγχο.

Επίσης, η Microsoft δεν υποστηρίζει την αφαίρεση μεταδεδομένων για δεδομένα που χρησιμοποιούν τη Διαχείριση Δικαιωμάτων Πληροφοριών (IRM)¹¹. Η εκτέλεση του Ελέγχου θα πρέπει να πραγματοποιηθεί προτού εφαρμοστεί η IRM.

5.5.5 VBA και ActiveX

Ο Έλεγχος Εγγράφου δεν αφαιρεί κώδικα ή σχόλια από λειτουργίες Visual Basic για Εφαρμογές (VBA) ούτε δεδομένα που σχετίζονται με έλεγχο ActiveX. Και στις δυο περιπτώσεις είναι αδύνατο για να τον Έλεγχο να καθορίσει εάν αφαιρεί κρίσιμα δεδομένα και επομένως τα αφήνει ως έχουν.

5.5.6 Άλλες πηγές μεταδεδομένων

Κάποια στοιχεία συνεργασίας ή ροής εργασίας ενσωματώνουν μια διεύθυνση e-mail σε ένα αρχείο ως μεταδεδομένα. Ο Έλεγχος Εγγράφου συνήθως αφαιρεί τις διευθύνσεις αυτού του τύπου, εκτός αν χρησιμοποιείται το στοιχείο αποστολή-προς-αναθεώρηση για την ενσωμάτωση της διεύθυνσης. Σε αυτή την περίπτωση ο Έλεγχος δεν αφαιρεί τη διεύθυνση επειδή υποθέτει ότι κάποιος θα θελήσει να στείλει το έγγραφο πίσω μετά την αναθεώρηση. Να ληφθεί υπόψη ότι ο Έλεγχος δεν αφαιρεί διευθύνσεις e-mail που έχουν προστεθεί στο περιεχόμενο ενός εγγράφου, φύλλου εργασίας ή παρουσίασης.

Επίσης, ο Έλεγχος Εγγράφου δεν αφαιρεί υπερσυνδέσμους (hyperlinks), εκτός κι αν περιέχονται σε κάποιου τύπου μεταδεδομένα που ο Έλεγχος αφαιρεί, όπως μια κεφαλίδα ή ένα υδατογράφημα.

Τα ονόματα αρχείων, οι διαδρομές αρχείων, τα ονόματα των προτύπων και οι διαδρομές των προτύπων μπορεί να περιέχουν μεταδεδομένα. Εν γένει, ο Έλεγχος Εγγράφου δεν αφαιρεί κανένα από αυτά από ένα αρχείο, καθώς είναι απαραίτητα.

¹¹ Η IRM (γνωστή κι ως e-DRM) είναι υποσύνολο των τεχνολογιών Διαχείρισης Ψηφιακών Δικαιωμάτων (DRM), οι οποίες προστατεύουν ευαίσθητες πληροφορίες από μη εγκεκριμένη πρόσβαση. [http://en.wikipedia.org/wiki/Digital_rights_management]

Ένας καλός κανόνας είναι να επιλέγει κανείς ονόματα προτύπων και αρχείων που να είναι γενικά και να μη χρησιμοποιεί συμβάσεις που να περιέχουν προσωπικές ή ιδιωτικές πληροφορίες.

Οι κώδικες πεδίων σε έγγραφα του Word μπορεί να είναι επίσης πηγή πληροφοριών, επειδή ο Έλεγχος αφαιρεί το περιεχόμενο αυτών αλλά δεν αφαιρεί τον κώδικα καθεαυτό. Για παράδειγμα αν προστεθεί ο κώδικας πεδίου συγγραφέα σε ένα έγγραφο, ο Έλεγχος θα αφαιρέσει το όνομα του συγγραφέα από τον κώδικα, αλλά θα κρατήσει τον κώδικα πεδίου συγγραφέα στο έγγραφο.

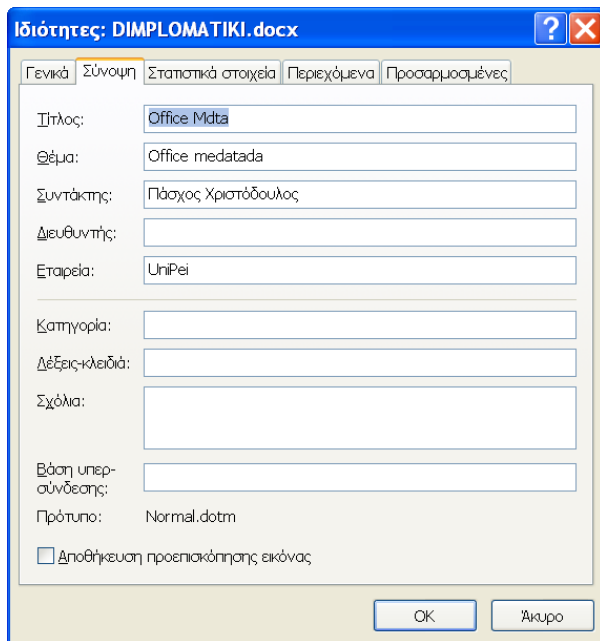
Κάποια άλλα πράγματα που ο Έλεγχος Εγγράφου δεν επιθεωρεί είναι τα: SmartArt, WordArt, σχήματα και γρήγορα τμήματα (quick parts). Ο Έλεγχος θεωρεί πως αυτά είναι μέρος του επιθυμητού περιεχομένου και δεν τα αφαιρεί, ούτε τις ετικέτες ούτε το κείμενο που προστίθεται σε αυτά.

Τέλος, ο Έλεγχος Εγγράφου δεν αφαιρεί προσαρμοσμένο κείμενο προτροπής (prompt text) από παρουσιάσεις PowerPoint. Μπορεί κανείς να προσθέσει προσαρμοσμένο κείμενο προτροπής σε ένα υπόδειγμα διαφανειών, αντικαθιστώντας το κείμενο που βλέπουν οι χρήστες όταν δημιουργούν νέες διαφάνειες. Θα πρέπει κανείς να τα αντικαθιστά ή να τα αφαιρεί εάν είναι προσαρμοσμένα και περιέχουν προσωπικές ή απόρρητες πληροφορίες, αλλιώς θα μπορεί να τις δει ο καθένας που ανοίγει την παρουσίαση και κοιτάζει το υπόδειγμα.

5.6 Επισκόπηση σύνθετων Ιδιοτήτων Ελέγχου Εγγράφου

Στην παρακάτω εικόνα βλέπουμε τις πληροφορίες που μας παρέχει η καρτέλα "Σύνοψη" στην προκειμένη περίπτωση, κάποιες από τις οποίες υπήρχαν και στις βασικές ιδιότητες που είδαμε προηγουμένως:

- Τίτλος
- Θέμα (κενό)
- Συντάκτης
- Διευθυντής (κενό)
- Εταιρία
- Κατηγορία (κενό)
- Λέξεις-κλειδιά (κενό)
- Σχόλια (κενό)
- Βάση υπερσύνδεσης (κενό)
- Πρότυπο
- Πλαίσιο επιλογής αποθήκευσης μικρογραφιών για όλα τα έγγραφα του Word (μη επιλεγμένο)



Εικόνα 6 Προβολή των Πληροφοριών Σύνοψης του Εγγράφου

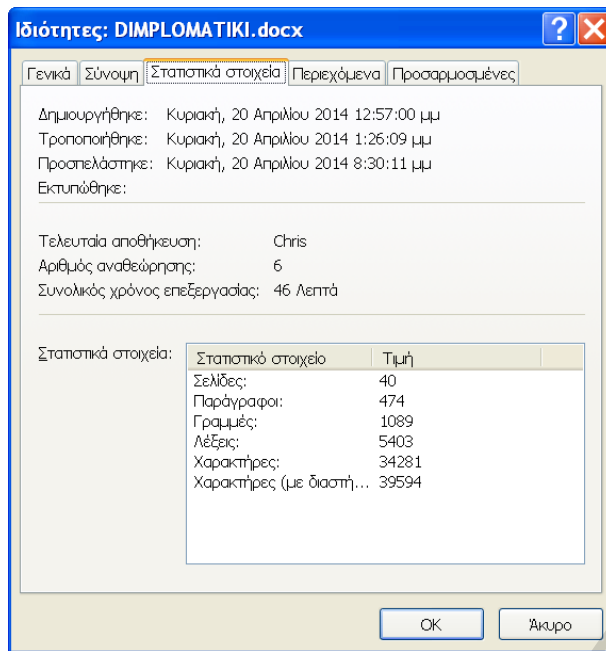
Στη συνέχεια προβάλλουμε τα περιεχόμενα τις καρτέλας "Στατιστικά Στοιχεία", τα οποία αποτελούν πληροφορίες του εγγράφου που δεν μπορούμε να τροποποιήσουμε εδώ:

- Ημερομηνία Δημιουργίας
- Ημερομηνία Τροποποίησης
- Ημερομηνία Προσπέλασης
- Ημερομηνία Εκτύπωσης
- Όνομα του ατόμου που πραγματοποίησε την τελευταία αποθήκευση
- Αριθμός αναθεωρήσεων
- Συνολικός χρόνος επεξεργασίας

Καθώς και τα στατιστικά στοιχεία:

- Αριθμός σελίδων
- Αριθμός παραγράφων
- Αριθμός γραμμών
- Αριθμός λέξεων
- Αριθμός χαρακτήρων

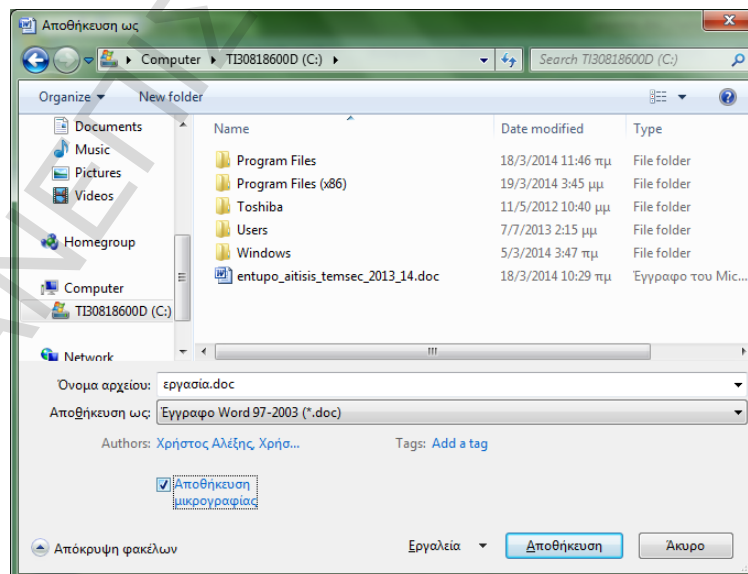
- Αριθμός χαρακτήρων (μαζί με τα διαστήματα)



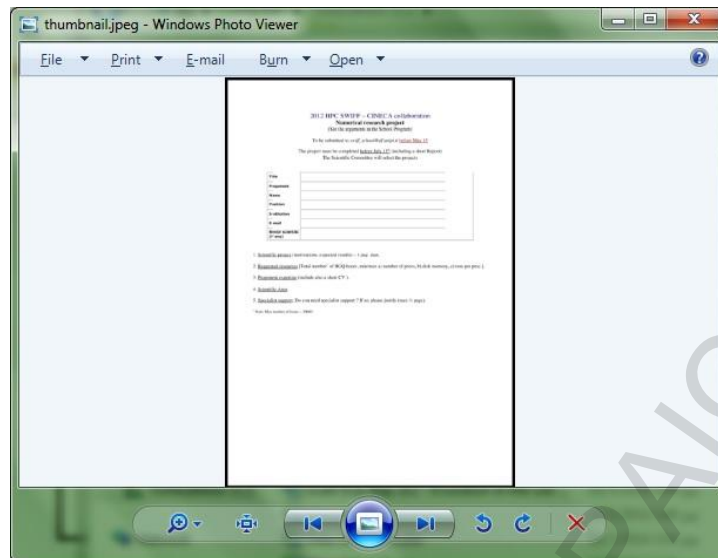
Εικόνα 7 Προβολή Πληροφοριών Στατιστικών Στοιχείων του Εγγράφου

5.7 Μικρογραφίες

Ένα ακόμη προνόμιο των νέων εγγράφων Office είναι οι μικρογραφίες (Thumbnails). Δεν είναι οι εικόνες που είναι ενσωματωμένες στα αρχεία κειμένου, αλλά συνήθως ένα στιγμιότυπο της πρώτης σελίδας του εγγράφου. Αποθηκεύεται ως μια εικόνα στον υπολογιστή σε μορφή .jpg. Παρόλα αυτά, τα Word 2007 και Excel 2007 δεν τις αποθηκεύουν από προεπιλογή. Στην εγκληματολογία έχουν σημασία καθώς η μικρογραφία θα πρέπει να ταιριάζει με το έγγραφο, επομένως αν δεν ταιριάζουν, είναι ένδειξη ότι το έγγραφο τροποποιήθηκε.



Εικόνα 8 Χειροκίνητη Επιλογή της Αποθήκευσης Μικρογραφίας



Εικόνα 9 Η Μικρογραφία ενός αρχείου .docx

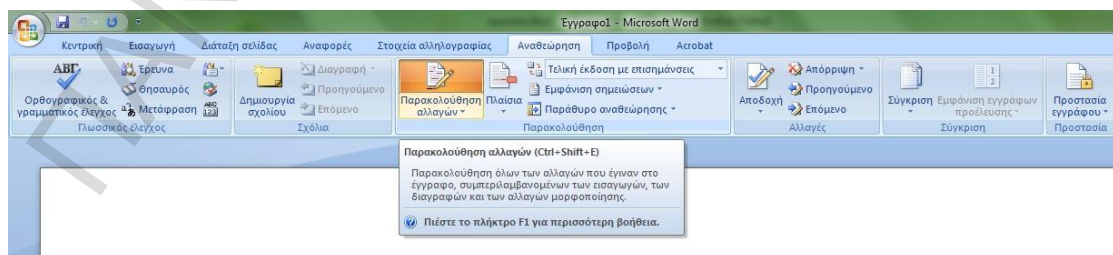
Βέβαια, ανεξαρτήτως αλλαγών το έγγραφο κρατά τα μοναδικά αναγνωριστικά, τα οποία είναι και ανεξάρτητα του μηχανήματος.

5.8 Χρονοσφραγίδες

Ένα ακόμη χαρακτηριστικό για εξέταση κατά την ανάλυση εγγράφων Office X είναι οι χρονοσφραγίδες (time-stamps). Εκτός από τις χρονοσφραγίδες του συστήματος αρχείων, τα έγγραφα Office X διατηρούν τις δικιές τους χρονοσφραγίδες ώστε να μπορούν να ξεχωρίζουν διαφορετικές εκδόσεις του ίδιου εγγράφου. Η ίδια δυνατότητα μπορεί να χρησιμοποιηθεί για να φανεί ποιος τροποποίησε το έγγραφο, πότε, καθώς και περισσότερες από μια ταυτόχρονες συνεδρίες επεξεργασίας, δείχνοντας έτσι αν ένα έγγραφο έχει παραποιηθεί.

5.9 Παρακολούθηση Αλλαγών

Η παρακολούθηση αλλαγών (Track Changes) χρησιμοποιείται σε περίπτωση που γίνονται αρκετές αναθεωρήσεις του ίδιου εγγράφου από έναν ή περισσότερους χρήστες. Προβάλλει όλες τις τροποποιήσεις που γίνονται σε ένα έγγραφο περιλαμβάνοντας εισαγωγές, διαγραφές, αλλαγές γραμμών και σχόλια.



Εικόνα 10 Ενεργοποίηση της Παρακολούθησης Αλλαγών στο Office 2007

Αν το χαρακτηριστικό Παρακολούθηση Αλλαγών έχει ενεργοποιηθεί, ακόμη και μετά το διαμοιρασμό του αρχείου (μέσω e-mail, στο τοπικό δίκτυο ή μέσω μιας φυσικής συσκευής), από προεπιλογή, το αρχείο ανοίγει στη λειτουργία

Παρακολούθησης Αλλαγών για να αποκαλύψει τις αλλαγές που έχουν γίνει. Επίσης, δείχνει μαζί με όλα τα σχόλια, αν έχουν προστεθεί από άλλους συγγραφείς/διορθωτές του εγγράφου μαζί με το όνομα του συγγραφέα.

5.10 Αποστολή για αναθεώρηση μέσω των εφαρμογών του MS Office

Το χαρακτηριστικό Αποστολή σε Παραλήπτη Mail για Αναθεώρηση επιτρέπει σε έγγραφα να αποσταλούν μέσω της προεπιλεγμένης εφαρμογής e-mail. Θα πρέπει να χρησιμοποιείται πολύ προσεκτικά καθώς περιέχει λεπτομέρειες ολόκληρου του εγγράφου.

Μπορεί να περιλαμβάνονται:

- πληροφορίες για το μηχάνημα από το οποίο στάλθηκε το έγγραφο
- όνομα χρήστη του συνδεδεμένου χρήστη
- διεύθυνση E-mail
- θέμα του E-mail

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ

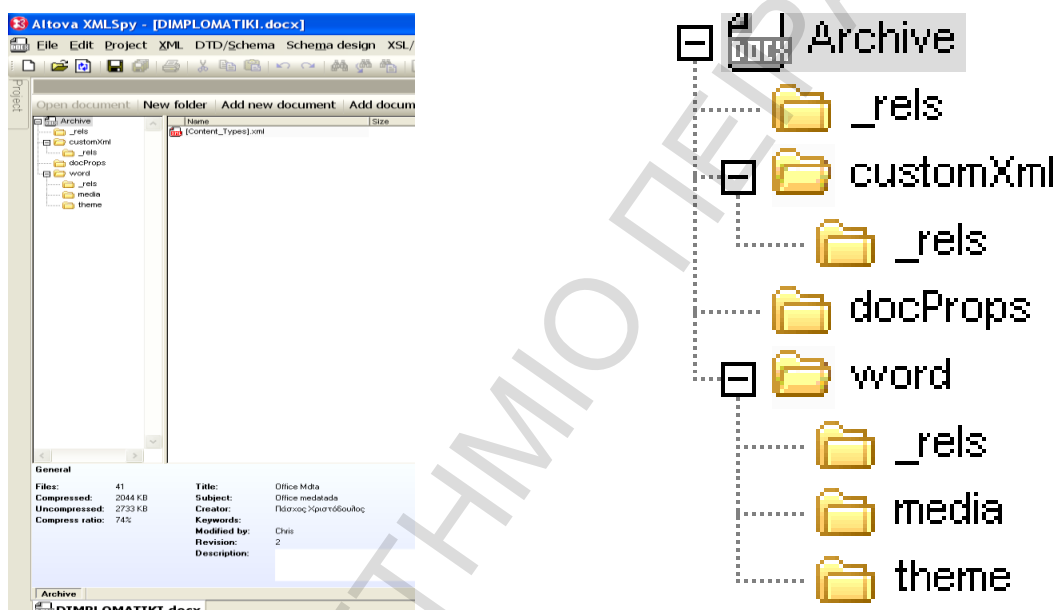
6 Άνοιγμα αρχείου .docx ως XML

6.1 Διάρθρωση OOXML

Θα προσπαθήσουμε τώρα να ανακτήσουμε μεταδεδομένα του εγγράφου χωρίς να το ανοίξουμε με το Word, αλλά αξιοποιώντας τις ιδιότητές του ως αρχείο XML.

Ακολουθούμε τις οδηγίες που αναφέρθηκαν κατά την περιγραφή αυτού του τύπου αρχείων.

Το έγγραφο έχοντας αλλάξει την επέκταση σε ZIP ή RAR ή ανοίγοντας το με κάποιο πρόγραμμα XML viewer¹² κατά το άνοιγμα περιέχει τους εξής φακέλους και αρχεία:



Εικόνα 11 Περιεχόμενα του Εγγράφου .docx κατά το άνοιγμα ως XML

- _rels (φάκελος)
- customXml (φάκελος)
- docProps (φάκελος)
- word (φάκελος)
- [Content_Types] (αρχείο XML)

¹² Χρησιμοποίησα την trial έκδοση του ALTOVA XMLSPY
<http://www.altova.com/xmlspy/office-open-xml.html>

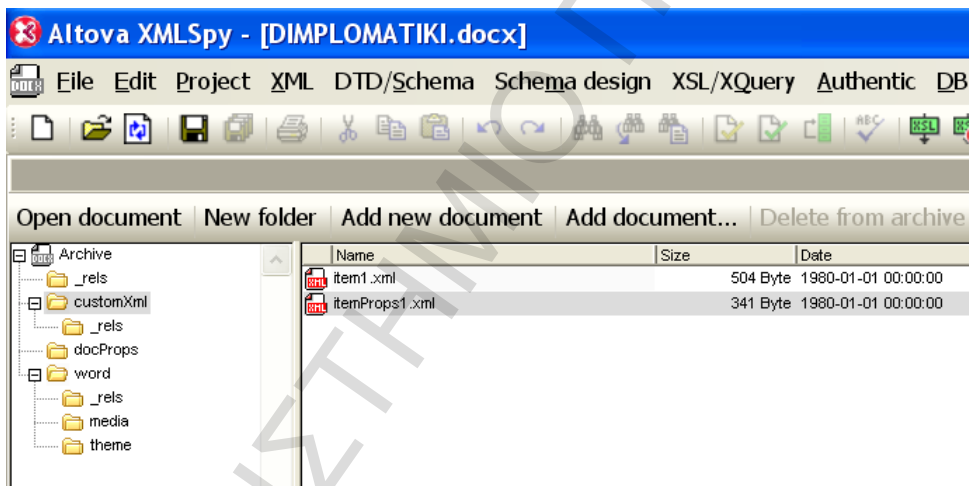
6.1.1 rels

Στο φάκελο `_rels` στο ανώτερο επίπεδο ενός πακέτου ZIP, το αρχείο που ονομάζεται `.rels` ορίζει σχέσεις μεταξύ των βασικών αρχείων του πακέτου ZIP και του σχήματος ανοιχτής μορφής XML του Office που εφαρμόζεται.

Και άλλοι φάκελοι μέσα σε ένα πακέτο ZIP μπορεί να περιέχουν δικούς τους φακέλους `_rels` για να ορίζουν σχέσεις εντός αυτού του φακέλου. Για παράδειγμα, ο κύριος φάκελος του εγγράφου περιέχει πάντα ένα φάκελο επιπέδου εγγράφου `_rels`.

Εκτός από τον ορισμό σχέσεων για τμήματα XML σε ένα πακέτο ZIP, τα αρχεία σχέσεων μπορεί επίσης να περιέχουν ορισμούς για σχέσεις με εξωτερικά αρχεία. Για παράδειγμα, αν το κείμενο του εγγράφου σας περιέχει υπερ-συνδέσεις, το αρχείο `document.xml.rels`¹³ θα περιέχει μια ξεχωριστή σχέση αναφοράς για κάθε υπερ-σύνδεση.

6.1.2 customXml



Εικόνα 12 Εντοπισμός μοναδικών αναγνωριστικών στο άνοιγμα ως XML

```
<?xml version="1.0" encoding="UTF-8" standalone="no"?>
<ds:datastoreItem ds:itemID="{C1433A9A-517F-4C2E-BB5E-00F91AB6B386}"14>
  xmlns:ds="http://schemas.openxmlformats.org/officeDocument/2006/customXml">
    <ds:schemaRefs>
      <ds:schemaRef
        ds:uri="http://schemas.openxmlformats.org/officeDocument/2006/bibliography"/>
    </ds:schemaRefs>
  </ds:datastoreItem>
```

Εικόνα 13 XML κώδικας του αρχείου `itemPros1.xml`

Στο αρχείο `itemPros1.xml`, το οποίο βρίσκεται στο φάκελο `customXml` και του οποίου τον κώδικα βλέπουμε παραπάνω, περιέχει ένα μοναδικό αναγνωριστικό ID, το

¹³ Όταν πρόκειται για έγγραφο του Word, το αρχείο που περιέχει ο φάκελος επιπέδου εγγράφου `_rels` ονομάζεται `document.xml.rels`.

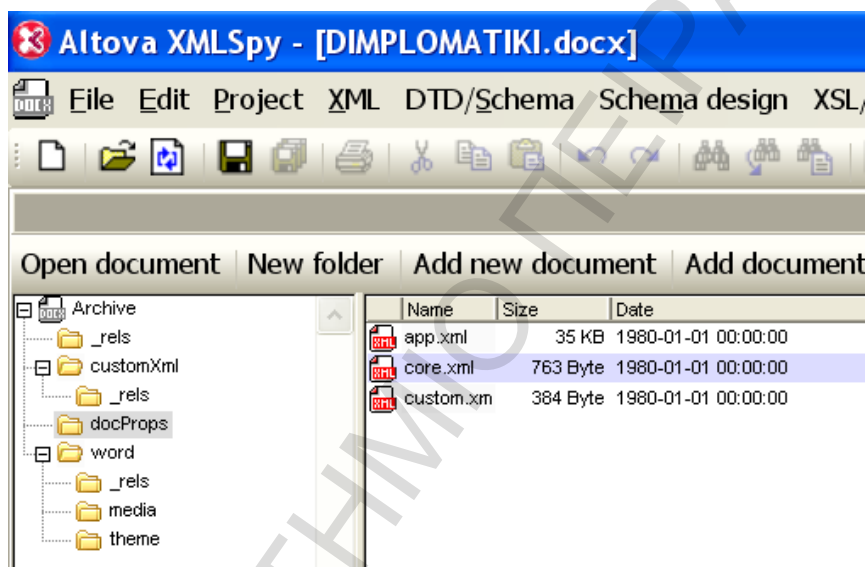
¹⁴ <http://msdn.microsoft.com/en-us/magazine/cc163478.aspx>

οποίο χρησιμοποιείται ως ξεχωριστή αναφορά του κώδικα που έχει προσθέσει ο χρήστης, σε σχέση με τον κώδικα της Microsoft ή το κυρίως έγγραφο. Στο παρελθόν τα μοναδικά αναγνωριστικά που περιέχονται στα έγγραφα του word έχουν χρησιμοποιηθεί σε συνδυασμό και με άλλα στοιχεία για τον εντοπισμό συγκεκριμένου χρήστη, όπως στη περίπτωση του εντοπισμού του συγγραφέα του ιού Melissa.

6.1.3 docProps

Ο φάκελος docProps περιέχει τα αρχεία:

- core.xml
- app.xml και
- custom.xml



Εικόνα 14 Περιεχόμενα του φακέλου docProps

Στο αρχείο custom.xml, βλέποντας τον κώδικα παρατηρούμε ότι περιέχει το custom πεδίο που έχω προσθέσει στις σύνθετες Ιδιότητες του Ελέγχου Εγγράφου.

```
<?xml version="1.0" encoding="UTF-8" standalone="yes"?>
<Properties
xmlns="http://schemas.openxmlformats.org/officeDocument/2006/custom-properties"
xmlns:vt="http://schemas.openxmlformats.org/officeDocument/2006/docPropsVTypes">
  <property fmtid="{D5CDD505-2E9C-101B-9397-08002B2CF9AE}"
pid="2" name="Αριθμός τηλεφώνου">
    <vt:r8>6977007492</vt:r8>
  </property>
</Properties>
```

Εικόνα 15 XML κώδικας του αρχείου custom.xml

Στο αρχείο core.xml, βλέποντας τον κώδικα παρατηρούμε ότι περιέχει Βασικές Ιδιότητες όπως:

```
<?xml version="1.0" encoding="UTF-8" standalone="yes"?>
<cp:coreProperties
xmlns:cp="http://schemas.openxmlformats.org/package/2006/metadata/core-
properties"
xmlns:dc="http://purl.org/dc/elements/1.1/"
xmlns:dcterms="http://purl.org/dc/terms/"
xmlns:dcmltype="http://purl.org/dc/dcmltype/"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance">
  <dc:title>Office Mtda</dc:title>
  <dc:subject>Office medatada</dc:subject>
  <dc:creator>Πάσχος Χριστόδουλος</dc:creator>
  <cp:keywords/>
  <cp:lastModifiedBy>Chris</cp:lastModifiedBy>
  <cp:revision>20</cp:revision>
  <dcterms:created
xsi:type="dcterms:W3CDTF">2014-04-
20T09:57:00Z</dcterms:created>
  <dcterms:modified
xsi:type="dcterms:W3CDTF">2014-04-
21T16:25:00Z</dcterms:modified>
</cp:coreProperties>
```

Εικόνα 16 XML κώδικας του αρχείου core.xml

- title (τίτλος)
- subject
- creator (δημιουργός-συντάκτης)
- keywords
- last modified by (τροποποιήθηκε τελευταία από-όνομα)
- revision (αναθεώρηση)
- last printed (τελευταία εκτύπωση-ημερομηνία-δεν έχει εκτυπωθεί ακόμα)
- created (δημιουργήθηκε-ημερομηνία)
- modified (τροποποιήθηκε-ημερομηνία)

Το αρχείο app.xml περιέχει Ιδιότητες Εφαρμογής, όπως σε αυτή την περίπτωση:

- Template (Πρότυπο)
- Total time (Συνολικός χρόνος)
- Pages (Σελίδες-αριθμός)
- Words (Λέξεις-αριθμός)
- Characters (Χαρακτήρες-αριθμός)
- Application (Εφαρμογή)
- DocSecurity (Ασφάλεια Εγγράφου)
- Lines (Γραμμές- αριθμός)
- Paragraphs (Παράγραφοι - αριθμός)
- Διάφορες Ιδιότητες Στυλ και Όνομα του Τίτλου (Title)
- Company (Εταιρία)
- Links up-to-date (Σύνδεσμοι έως τώρα)
- Characters with spaces (Χαρακτήρες με διαστήματα)

- SharedDoc (Διαμοιρασμός του εγγράφου)
- Hyperlinks Changed (Αλλαγμένοι Υπερσύνδεσμοι)
- AppVersion (Έκδοση της Εφαρμογής- αριθμός)

```

<?xml version="1.0" encoding="UTF-8" standalone="yes"?>
<Properties
xmlns="http://schemas.openxmlformats.org/officeDocument/2006/extended-
properties"
xmlns:vt="http://schemas.openxmlformats.org/officeDocument/2006/docPropsVType
s">
  <Template>Normal.dotm</Template>
  <TotalTime>671</TotalTime>
  <Pages>49</Pages>
  <Words>8617</Words>
  <Characters>46535</Characters>
  <Application>Microsoft Office Word</Application>
  <DocSecurity>0</DocSecurity>
  <Lines>387</Lines>
  <Paragraphs>110</Paragraphs>
  <ScaleCrop>>false</ScaleCrop>
  <HeadingPairs>
    <vt:vector size="4" baseType="variant">
      <vt:variant>
        <vt:lpstr>Τίτλος</vt:lpstr>
      </vt:variant>
      <vt:variant>
        <vt:i4>1</vt:i4>
      </vt:variant>
      <vt:variant>
        <vt:lpstr>Title</vt:lpstr>
      </vt:variant>
      <vt:variant>
        <vt:i4>1</vt:i4>
      </vt:variant>
    </vt:vector>
  </HeadingPairs>
  <TitlesOfParts>
    <vt:vector size="2" baseType="lpstr">
      <vt:lpstr>Office Mda</vt:lpstr>
      <vt:lpstr/>
    </vt:vector>
  </TitlesOfParts>
  <Company>UniPei</Company>
  <LinksUpToDate>>false</LinksUpToDate>
  <CharactersWithSpaces>55042</CharactersWithSpaces>
  <SharedDoc>>false</SharedDoc>
  <HLinks> <!-- ακολουθούν τα link που περιέχονται στο έγγραφο -->
    <vt:vector size="738" baseType="variant">

```

Εικόνα 17 XML κώδικας του αρχείου app.xml

Παρατηρούμε ότι ανοίγοντας αυτά τα δύο αρχεία βρίσκουμε μεταδεδομένα του εγγράφου που βρήκαμε και εξερευνώντας τις ιδιότητες του MS Word.

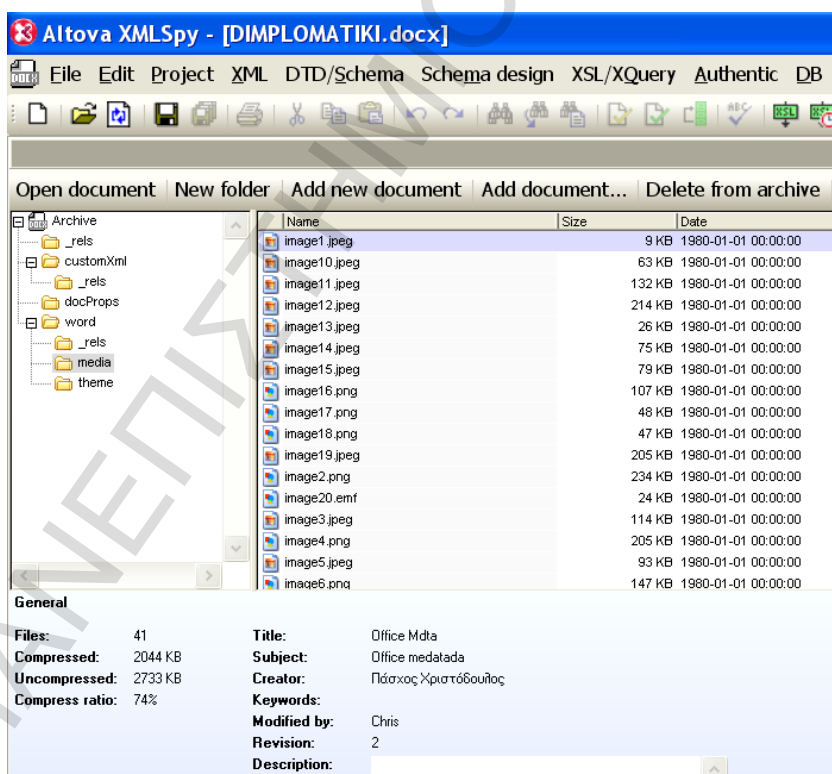
6.1.4 word

Ο κύριος φάκελος του εγγράφου σε ένα πακέτο ZIP - OOXML, περιέχει αρκετά μέρη XML και επιπρόσθετους φακέλους.

Ο φάκελος word περιέχει το δικό του φάκελο _rels, ο οποίος περιέχει το αρχείο document.xml.rels. Το συγκεκριμένο αρχείο περιέχει τις σχέσεις για τα μέρη XML (όπως για τα αρχεία document.xml και fontTable.xml). Ο φάκελος επίσης περιέχει επιπρόσθετα αρχεία .rels, ανάλογα με το περιεχόμενο του συγκεκριμένου εγγράφου.

Ο υποφάκελος Theme, περιέχει τον ορισμό για το θέμα εγγράφου που χρησιμοποιείται στο έγγραφο, συμπεριλαμβανομένων των χρωμάτων θέματος, των γραμματοσειρών θέματος και των εφέ θέματος.

Τα μέρη XML ποικίλλουν ανάλογα με τον τύπο του αρχείου. Τα έγγραφα που δημιουργούνται έχουν πάντα τα τμήματα XML που αναφέρονται παραπάνω, ενώ μπορεί να περιέχουν και επιπρόσθετα τμήματα ανάλογα με το περιεχόμενο του εγγράφου (όπως π.χ., ένα τμήμα για κάθε κεφαλίδα και υποσέλιδο ή ένα για τα σχόλια).



Εικόνα 18 Πολυμέσα αρχείου word κατά το άνοιγμα ως XML

Ο φάκελος word μπορεί να περιέχει και επιπρόσθετους φακέλους, ανάλογα με το περιεχόμενο του αρχείου. Για παράδειγμα, αν το αρχείο περιέχει εικόνες όπως στο εικόνα που φαίνεται παραπάνω, υπάρχει ένας φάκελος πολυμέσα που περιέχει κάθε εικόνα του εγγράφου αποθηκευμένη σε ένα ξεχωριστό αρχείο εικόνας.

Ένα θέμα που αντιμετωπίζει ο ερευνητής σε αυτό το σημείο, είναι ότι τα metadata, των αρχείων που περιέχονται στο έγγραφο, όπως οι φωτογραφίες στο παράδειγμά μας, δεν είναι σωστά.

6.1.5 Content_Types

Το αρχείο [Content_Types].xml προσδιορίζει τύπους βασικού περιεχομένου στο έγγραφο, όπως το κυρίως σώμα, τα στυλ, τις ρυθμίσεις και τις ιδιότητες αρχείου του εγγράφου. Όπως προαναφέραμε, αυτό το αρχείο προσδιορίζει επίσης τύπους αρχείων που περιλαμβάνονται στο πακέτο ZIP.

Κάθε τύπος αρχείου στο πακέτο ZIP, συμπεριλαμβανομένων και των αρχείων .xml, .rels και οποιωνδήποτε άλλων αρχείων (όπως το αρχείο εικόνας .png που αναφέρεται στην εικόνα), προσδιορίζεται με μια αναφορά Προεπιλεγμένης επέκτασης.

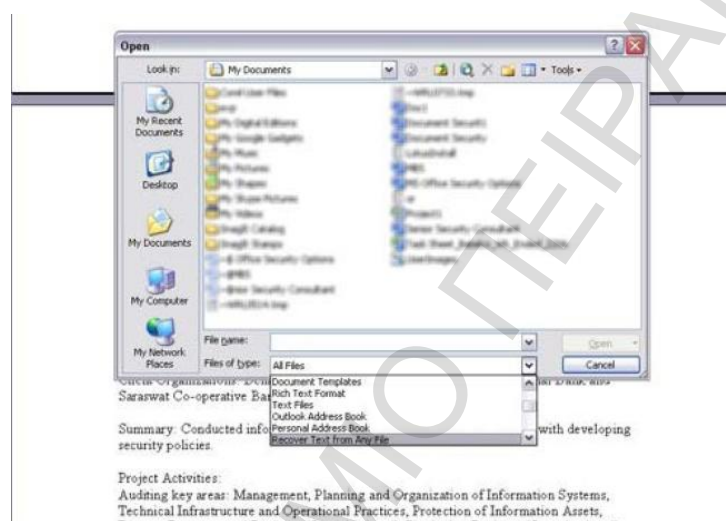
Οι βασικοί τύποι περιεχομένου (συνήθως εκείνοι για τους οποίους υπάρχει ξεχωριστό τμήμα XML στο πακέτο ZIP) υποδεικνύονται με μια αναφορά Παράκαμψης ονόματος τμήματος. Η "Παράκαμψη ονόματος τμήματος" εκφράζεται ως η διαδρομή προς το τμήμα XML εντός του πακέτου ZIP (όπως/word/document.xml).

Τόσο η αναφορά "Προεπιλεγμένης επέκτασης" όσο και η αναφορά "Παράκαμψης ονόματος τμήματος" παρέχουν έναν ορισμό Τύπου περιεχομένου, όπως βλέπετε στα παραδείγματα της εικόνας, η οποία προέρχεται από το σχήμα που εφαρμόζεται.

7 Μεταδεδομένα σε Παλαιότερες Εκδόσεις του Office

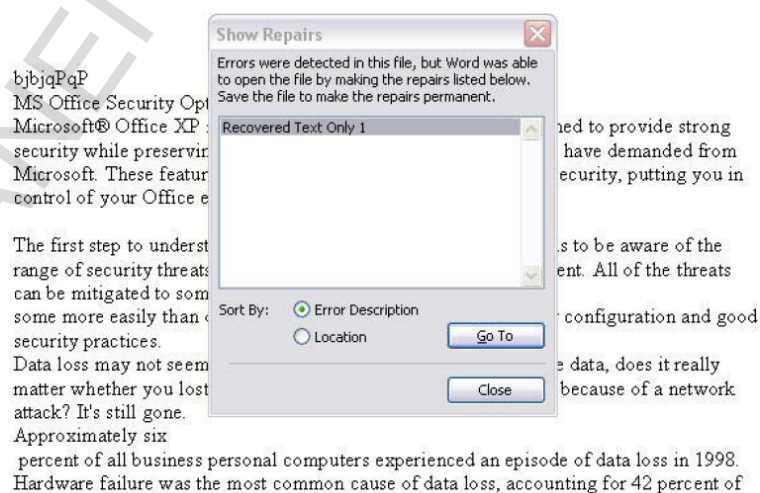
Από προεπιλογή τα έγγραφα του MS Office 2003 περιέχουν κρυμμένα δεδομένα, η ανακάλυψη των οποίων δεν είναι τόσο δύσκολη. Για παράδειγμα ένα χαρακτηριστικό του Word επιτρέπει το άνοιγμα κατεστραμμένου εγγράφου με προβολή του κειμένου χωρίς μορφοποίηση. Μπορεί επίσης να χρησιμοποιηθεί για την προβολή ορισμένων μεταδεδομένων που σχετίζονται με ένα έγγραφο, με τα εξής βήματα:

Αρχείο → Άνοιγμα → λίστα Αρχεία Τύπου → "Ανάκτηση Κειμένου από Οποιοδήποτε Αρχείο" → εντοπισμός ενός αρχείου Word *.doc → "Άνοιγμα".



Εικόνα 19 Ανάκτηση Κειμένου από Οποιοδήποτε Αρχείο

Το έγγραφο ανοίγει χωρίς καμία μορφοποίηση. Μετά την κύλιση μέσα στο έγγραφο, μπορεί κανείς να δει πληροφορία, τέτοια όπως το όνομα του συγγραφέα του εγγράφου, τη διαδρομή του αποθηκευμένου εγγράφου κ.ο.κ. Η πληροφορία που φαίνεται μπορεί να μην είναι στο περιεχόμενο. Θα πρέπει να είναι κανείς προσεκτικός για να διακρίνει ποια πληροφορία από αυτή που βλέπει είναι μέρος του κειμένου του εγγράφου ή των μεταδεδομένων που προστέθηκαν στο έγγραφο.

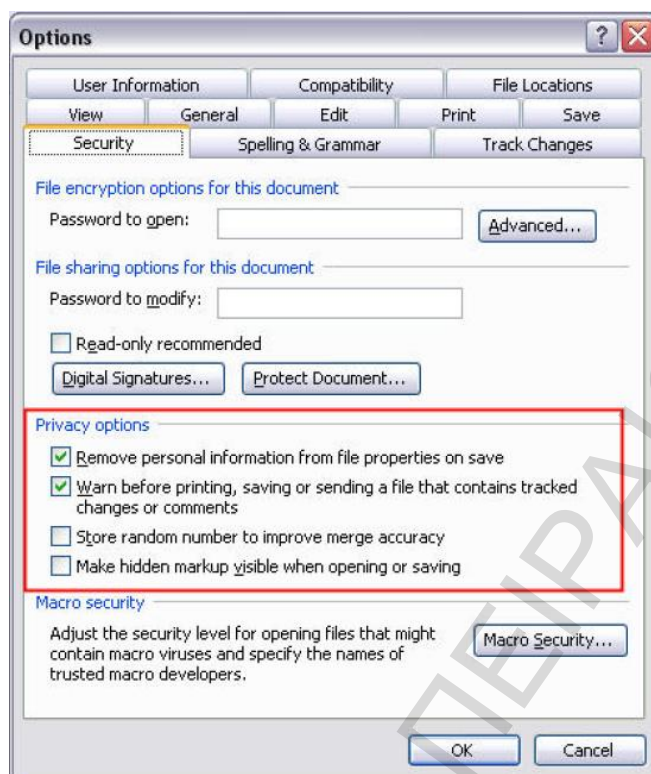


Εικόνα 20 Πλαίσιο διαλόγου μετά την ανάκτηση αρχείου

Times New Roman
Symbol
Symbol
Tahoma
Tahoma
MS Office Security Options
MS Office Security Options
User 2
User 2
MS Office Security Options
User 2
Normal
Microsoft Office Word
MS Office Security Options
Root Entry
1Table
1Table
WordDocument
WordDocument
SummaryInformation
SummaryInformation
DocumentSummaryInformation
DocumentSummaryInformation
CompObj
CompObj
Microsoft Office Word Document
MSWordDoc
Word.Document.8

Εικόνα 21 Δείγμα περιεχομένων ανακτηθέντος αρχείου

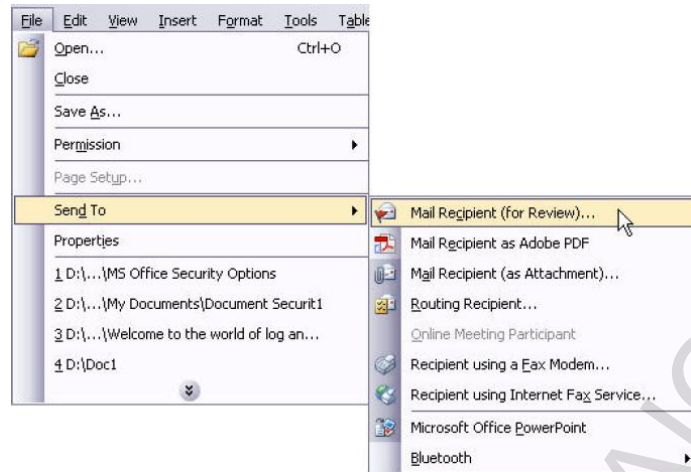
Αν κανείς χρησιμοποιεί την "Παρακολούθηση Αλλαγών" από το μενού "Εργαλεία", "Εκδόσεις" από το μενού "Αρχείο", ή την επιλογή "Να επιτρέπεται η γρήγορη αποθήκευση" από την καρτέλα "Αποθήκευση του πλαισίου διαλόγου" "Επιλογές", διαθέσιμο στο μενού "Εργαλεία", θα πρέπει να κοιτάξει να αφαιρέσει όποια κρυμμένη ή διαγραμμένη πληροφορία που μπορεί να παραμένει στο έγγραφο.



Εικόνα 22 Απομάκρυνση Προσωπικών Πληροφοριών από Ιδιότητες Αρχείου

Το Microsoft Office 2003 έχει επίσης την ικανότητα να διενεργεί ταχείες αποθηκεύσεις (Fast saves), σε αντιστοιχία με την Αυτόματη Αποθήκευση των μεταγενέστερων εκδόσεων. Αν ο χρήστης συνθέσει ένα έγγραφο, ενώ η "Ταχεία Αποθήκευση" είναι ενεργοποιημένη, και διαγράψει κάποιο κείμενο, αυτό το διαγραμμένο κείμενο είναι πιθανόν να παραμείνει με το έγγραφο σε όλη τη διάρκεια ζωής του. Κάποιο κείμενο απλά δε διαγράφεται ποτέ.

Αν το έγγραφο έχει σταλεί μέσω του Outlook, όταν ο παραλήπτης ανοίξει το έγγραφο και δει τις ιδιότητες του εγγράφου, προβάλλονται εγγραφές όπως `_TentativeReviewCycleID` και `_ReviewCycleID`, `_EmailSubject`, `_AuthorEmail`, και `_AuthorEmailDisplayName`. Οι πληροφορίες αυτές αποθηκεύονται επίσης στο σύστημα του παραλήπτη σε ένα αρχείο που ονομάζεται 'Adhoc.rcd' ή 'Review.rcd' (ανάλογα με την έκδοση του MS Office που χρησιμοποιείται, π.χ. 'Review.rcd' είναι στην περίπτωση του Office 2003). Συνήθως βρίσκεται μέσα στο φάκελο Έγγραφα και Ρυθμίσεις Χρήστη > Δεδομένα Εφαρμογών \ Microsoft \ Office.



Εικόνα 23 Αποστολή Εγγράφου για Αναθεώρηση

8 Μεταδεδομένα σε Αρχεία του OpenOffice.org

Το OpenOffice.org σκιαγραφεί στις προδιαγραφές της μορφής των XML αρχείων του τη δυνατότητα αποθήκευσης πολλών μεταδεδομένων μέσα στα έγγραφα, λογιστικά φύλλα και στις παρουσιάσεις του:

- Πηγή
- Τίτλος
- Περιγραφή
- Θέμα
- Λέξεις κλειδιά
- Αρχικός δημιουργός
- Δημιουργός
- Εκτυπώθηκε από
- Ημερομηνία και ώρα δημιουργίας
- Ημερομηνία και ώρα τροποποίησης
- Ημερομηνία και ώρα εκτύπωσης
- Πρότυπο εγγράφου
- Γλώσσα
- Μεταδεδομένα ορισμένα από το χρήστη
- Στατιστικά εγγράφου

```
<office:meta>
<dc:title>Title</dc:title>
<dc:description>Comment</dc:description>
<dc:language>en-US</dc:language>
<meta:initial-creator>Name</meta:initial-
creator>
<meta:creation-date>X</meta:creation-date>
<dc:creator>Name</dc:creator>
<meta:generator>NeoOffice</meta:generator>
<dc:creator>Name</dc:creator>
<meta:keywords>
<meta:keyword>First</meta:keyword>
<meta:keyword>Second</meta:keyword>
</meta:keywords>
<dc:date>X:</dc:date>
<dc:subject>Subject</dc:subject>
<meta:printed-by>Name</meta:printed-by>
<meta:print-date>X</meta:print-date>
<meta:duration-time>X</meta:editing-
duration>
<meta:editing-cycles>4</meta:editing-cycles>
<meta:editing-duration>X</meta:editing-
duration>
</office:meta>
```

Εικόνα 24 Λείψμα μεταδεδομένων σε OpenOffice.org Writer έγγραφο

Σύμφωνα με το έγγραφο των προδιαγραφών, όλες οι παραπάνω κατηγορίες αποθηκεύονται ή έχουν τη δυνατότητα να αποθηκευτούν με κάθε αρχείο που δημιουργείται στο OpenOffice.org. Αυτή είναι σε μεγάλο βαθμό πληροφορία του είδους που αποθηκεύεται κι από το Microsoft Office, παρόλα αυτά το OpenOffice.org δεν κάνει υποθέσεις για το τι προσωπική πληροφορία θα ήθελε να συσχετίσει ο χρήστης με το αρχείο. Για παράδειγμα αν ένας χρήστης δεν εισάγει ένα όνομα στο πλαίσιο πληροφοριών χρήστη της εφαρμογής, το πρόγραμμα δε θα προσπαθήσει να λάβει την πληροφορία εγγραφής από το λειτουργικό σύστημα προκειμένου να ανακαλύψει το όνομα ή τα αρχικά του χρήστη.

Όπως και το Microsoft Office, το OpenOffice.org έχει τη δυνατότητα να αποθηκεύει πληροφορία για να διευκολύνει τη χρήση της λειτουργίας "Παρακολούθηση" και τα χαρακτηριστικά σχολιασμού. Το OpenOffice.org παρέχει ευκολότερη πρόσβαση στα μεταδεδομένα ενός αρχείου κάτι που διευκολύνει την ανεύρεσή τους από το χρήστη και τον εξεταστή, δεδομένου ότι αυτοί ξέρουν για το τι ψάχνουν. Το OpenOffice.org αποθηκεύει τα μεταδεδομένα που σχετίζονται με το συνολικό έγγραφο, φύλλο εργασίας ή την παρουσίαση σε ένα ξεχωριστό απλό αρχείο κειμένου για ευκολότερη πρόσβαση.

9 Εμπορικά εργαλεία

Σε αυτή την ενότητα, αναφέρονται ενδεικτικά κάποια από τα εργαλεία που πραγματοποιούν αυτοματοποιημένη εύρεση και κατάργηση δεδομένων από έγγραφα, μεταξύ άλλων αρχείων. Πολλά από αυτά τα προγράμματα προσφέρουν πολύ περισσότερες δυνατότητες από την ανάκτηση και απομάκρυνση μεταδεδομένων.

- Encase - Guidance Software
<http://www.guidancesoftware.com/>

Το Encase είναι μια πολύ δημοφιλής σουίτα προϊόντων ψηφιακών πειστηρίων. Κυκλοφορεί σε διάφορες μορφές, σχεδιασμένες για διάφορες χρήσεις: εγκληματολογικές, κυβερνοασφάλειας και ηλεκτρονικών αποκαλύψεων (EnCase Forensic, EnCase Cybersecurity, EnCase eDiscovery, EnCase Portable).

Αν και πραγματοποιεί αυτοματοποιημένα τις διάφορες λειτουργίες, ακόμα και κάποιος εξειδικευμένος ερευνητής χρειάζεται να παρακολουθήσει σεμινάρια για την πλήρη αξιοποίησή του. Τα δεδομένα που ανακύπτουν από το EnCase έχουν χρησιμοποιηθεί επιτυχώς σε διάφορα δικαστικά συστήματα σε όλο τον κόσμο.

Το EnCase έχει πρόσβαση σε δεδομένα που ένας χρήστης δε θα μπορούσε να έχει μέσω του λειτουργικού του συστήματος. Επιπρόσθετα, παρέχει τη λειτουργία προβολής ατομικών τμημάτων σύνθετων τύπων αρχείων, όπως αρχεία μητρώου, αρχεία OLE, αρχεία e-mail του MS Outlook, Windows Thumbs.db, και αρχεία Macintosh PAX.

- Oracle Outside In
<http://www.oracle.com/technetwork/middleware/webcenter/content/oit-all-085236.html>

Η Outside In Technology είναι μια σουίτα εργαλείων ανάπτυξης λογισμικού (SDKs) που παρέχει στους προγραμματιστές μια κατανοητή λύση για την εξαγωγή, κανονικοποίηση, αφαίρεση, μετατροπή και προβολή των περιεχομένων 600 μη δομημένων τύπων αρχείων.

Κάθε SDK μέσα στη σουίτα, βελτιστοποιείται για την επίλυση ενός συγκεκριμένου προβλήματος αλλά είναι πολύ ευέλικτα και διαλειτουργικά. Οι προγραμματιστές μπορούν εύκολα να εφαρμόσουν όποιο συνδυασμό των Outside In SDKs για την παροχή ακριβώς της σωστής λειτουργίας στην εφαρμογή τους, ελαχιστοποιώντας την προσπάθεια ενσωμάτωσης και τα αποτυπώματα του κώδικα. Τα SDKs προσφέρουν ευρύ φάσμα επιλογών για να δώσουν προγραμματιστικό έλεγχο στον προγραμματιστή για τη ροή της εργασίας και το αποτέλεσμα. Περιλαμβάνονται αναλυτική τεκμηρίωση και δείγματα εφαρμογών με πηγαίο κώδικα για την περαιτέρω επιτάχυνση της εφαρμογής.

Η Outside In Content Access εξάγει κείμενο και μεταδεδομένα από σχεδόν 500 τύπους αρχείων και μεταφράζει αυτόματα το κείμενο και τις ιδιότητες από πολλαπλές πιθανές κωδικοποιήσεις σε μια μοναδική κωδικοποίηση που καθορίζεται από τον προγραμματιστή. Είναι βελτιστοποιημένη για απόδοση, παρέχοντας δεδομένα διαδραστικά στην εφαρμογή τερματικού στη μνήμη, καθώς επεξεργάζεται το αρχείο εισόδου. Χρησιμοποιείται ευρέως στην έρευνα και εγκληματολογική εξέταση εφαρμογών. Είναι διαθέσιμη για περιβάλλοντα Windows, Unix, Linux και Solaris.

Η Clean Content εξάγει και καθαρίζει κείμενο, μεταδεδομένα και κρυφή πληροφορία από Word, Excel, PowerPoint και PDF. Η έκδοση 2012.1 επιτρέπει την εξαγωγή από επιπλέον 600 τύπους.

- PayneGroup Metadata Assistant
<http://www.thepaynegrup.com/products/metadata/>

Το Metadata Assistant παρέχει τις επιλογές για απομάκρυνση μεταδεδομένων και για επιλογή αρχείων σε μια μοναδική διεπιφάνεια για όλους τους τύπους αρχείων, προσφέροντας μεγαλύτερη ευελιξία στη λήψη αποφάσεων για την αφαίρεση συγκεκριμένων μεταδεδομένων από πολλαπλά αρχεία.

Ενσωματώνεται στα Microsoft Word, Excel και PowerPoint 2013, 2010 και 2007, καθώς και για E-mail με τα Outlook 2013, 2010, 2007, μεταξύ άλλων προγραμμάτων. Ανιχνεύει αυτόματα όλους τους τύπους αρχείων που υποστηρίζει, όταν αυτά είναι συνημμένα σε μήνυμα e-mail. Υποστηρίζει μετατροπές σε PDF όπως οι Adobe Acrobat 9 και X και παρέχει αποτελέσματα ανάλυσης σε μορφοποιήσεις RTF και XML.

- BEC MetaReveal
<http://www.beclegal.com/products/metareveal>

Το BEC MetaReveal θα απομακρύνει όποια μεταδεδομένα θέλει κανείς να μη συμπεριλάβει στα έγγραφά του. Μπορεί να αφαιρέσει όλα τα μεταδεδομένα από το έγγραφο Office ή να γίνει προβολή και επιλογή των μεταδεδομένων προς διατήρηση ή απομάκρυνση. Είναι καλό που συνδέεται με το Microsoft Outlook και προτρέπει για αφαίρεση μεταδεδομένων κατά την πίεση του κουμπιού Αποστολή, κάνοντάς αδύνατο να ξεχάσει κάποιος τον έλεγχο των μεταδεδομένων ενός αρχείου του Office πριν αποσταλεί με e-mail.

- Esquire Innovations iScrub 7 EP1
<http://esqinc.com/products/iscrub/>

Αυτό το εργαλείο όχι μόνο εκκαθαρίζει τα έγγραφα και διαχειρίζεται τα μεταδεδομένα αλλά επίσης επιτρέπει τη μετατροπή των εγγράφων σε PDF μετά την εκκαθάριση. Βρίσκει οτιδήποτε, από πληροφορίες παλαιότερου συγγραφέα του αρχείου έως παρακολουθούμενες αλλαγές, διαγραμμένο κείμενο και όποια άλλα μεταδεδομένα θέλει κανείς να αποκρύψει από τον παραλήπτη του αρχείου είτε μέσω email ή όποια άλλη μέθοδο διαμοιρασμού των αρχείων Office.

- Belkasoft Evidence Center
<http://belkasoft.com/>

Εξάγει μεταδεδομένα από διάφορα αρχεία Microsoft Office (97-2003 και 2007-2013 τύπους), όπως και έγγραφα Open Office. Εκτός αυτού, μπορεί να εξάγει απλά κείμενα (συνδυάζοντας όλα τα κείμενα από σελίδες XLS/XLSX/ODS και διαφάνειες PPT/PPTX/ODP) και ενσωματωμένα αντικείμενα. Για εικόνες, ενσωματωμένες σε ένα έγγραφο, το εργαλείο μπορεί να τις οπτικοποιήσει όλες στη διεπιφάνεια του χρήστη.

- Forensic Innovations FI Tools
<http://forensicinnovations.com/>

Η τεχνολογία Forensic Investigator έχει συγχωνευτεί με την Αναζήτηση των Windows και την εντολή Καταλόγου (Directory) του DOS κι επιτρέπει την αναζήτηση αρχείων ανάλογα με τον τύπο, το περιεχόμενο, το λειτουργικό σύστημα ή πλατφόρμα, τη μέθοδο αποθήκευσης και τα προσδιοριστικά τους μεταξύ άλλων. Υποστηρίζει πάνω από 100 τύπους αρχείων, υποσχόμενο μεγαλύτερη ακρίβεια από τα ανταγωνιστικά προγράμματα. Μπορεί να προσπελάσει αρχεία ακόμη κι αν είναι κλειδωμένα, απόρρητα ή κωδικοποιημένα, υπολογίζει κώδικες κατακερματισμού όπως οι MD5 και SHA-1 κι έχει πελάτες όπως την κυβέρνηση του Ηνωμένου Βασιλείου και τους κολοσσούς Ernst&Young και KPMG.

- Metadata Touch
<http://digitalconfidence.com/MetadataTouch.html>

Το metadata Touch έχει τη δυνατότητα να απεικονίζει και να τροποποιεί metadata για αρχεία Microsoft Office Word, Excel, PowerPoint, OpenOffice, JPEG, JPEG 2000, AVI, MP3, MP4, F4V, WAVE, AIFF, PNG, SVG, APE, MPC, OFR, WV, TAK, and XMP files.

Είναι διαθέσιμο σε δυο εμπορικές εκδόσεις ενώ υπάρχει και δοκιμαστική έκδοση 14 ημερών. Την trial έκδοση του metadata Touch θα αναλύσουμε λίγο παρακάτω.

10 Ελεύθερης χρήσης ή Ανοιχτού κώδικα εργαλεία

- Brightfort Doc Scrubber
<http://www.brightfort.com/>

Δωρεάν πρόγραμμα για προσωπική, εκπαιδευτική ή εταιρική χρήση. Αναλύει έγγραφα του Word και ανακαλύπτει κρυφά ή πιθανώς εμπιστευτικά δεδομένα που μπορεί να περιέχουν. Υποστηρίζει την εξέταση πολλαπλών αρχείων σε ένα φάκελο ταυτόχρονα λειτουργεί όμως μόνο με έγγραφα Word 97, 2000, XP και 2003.

- Libextractor και Extract
<http://gnunet.org/libextractor/>

Το εργαλείο ανοιχτού κώδικα και γραμμής εντολών Libextractor είναι η βιβλιοθήκη που χρησιμοποιείται για την εξαγωγή μεταδεδομένων από αρκετά διαφορετικούς τύπους αρχείων και το εργαλείο Extract είναι το πρόγραμμα που δουλεύει στη γραμμή εντολών για την εξαγωγή. Και τα δυο είναι διαθέσιμα για πλατφόρμες Linux and Windows δωρεάν και διαμοιραζόμενα με την γενική άδεια δημόσιας χρήσης GNU.

- Document Metadata Cleaner - Pointstone Software
<http://www.pointstone.com/products/metadata-cleaner/>

Αυτό το εργαλείο επιτρέπει την ανάλυση όλων των εγγράφων του Office (doc, xls, ppt) για την ανακάλυψη και μετέπειτα διαγραφή των μεταδεδομένων που δεν επιθυμεί κανείς να διαμοιραστεί. Είναι δωρεάν για εταιρική ή προσωπική χρήση και είναι δυνατό να εκκαθαρίσει ένα ή πολλαπλά έγγραφα ταυτόχρονα. Δεν υποστηρίζει εκδόσεις 2007 και μεταγενέστερες.

- Metadata Analyzer
<http://www.smartpctools.com/metadata/>

Το Metadata Analyzer είναι μια δωρεάν εφαρμογή για χρήση σε Windows. Ο μοναδικός σκοπός του είναι να αναφέρει τα μεταδεδομένα που βρίσκονται σε έγγραφα του Microsoft Word, παρουσιάσεις του PowerPoint, και φύλλα εργασίας του Excel. Όταν ανοίγει προσφέρει μια διεπιφάνεια χρήστη που αρχικά ανιχνεύει και αναφέρει τις ενσωματωμένες ιδιότητες του Office: Όνομα Χρήστη, Αρχικά Χρήστη και Εταιρία. Ο χρήστης έχει την ευχέρεια να καταδείξει ένα αρχείο στο πρόγραμμα για ανάλυση ή να σύρει το εικονίδιο του αντίστοιχου εγγράφου Word, PowerPoint, ή Excel πάνω στο πρόγραμμα (click and drop). Το πρόγραμμα προσφέρει ευανάγνωστη αναφορά, αλλά όχι τη δυνατότητα αποθήκευσης.

11 Μελέτη Περίπτωσης: Εξαγωγή Δεδομένων από τα έγγραφα του MS Word

Θα εξετάσουμε βήμα προς βήμα ένα αρχείο (*.docx), , για την ύπαρξη μεταδεδομένων.

Θα προσπαθήσουμε να επαληθεύσουμε ενδεικτικά την ύπαρξη πολλών από τους τύπους μεταδεδομένων που έχουμε ήδη αναφέρει, χειροκίνητα αλλά και με τη χρήση προγραμμάτων ελεύθερης χρήσης.

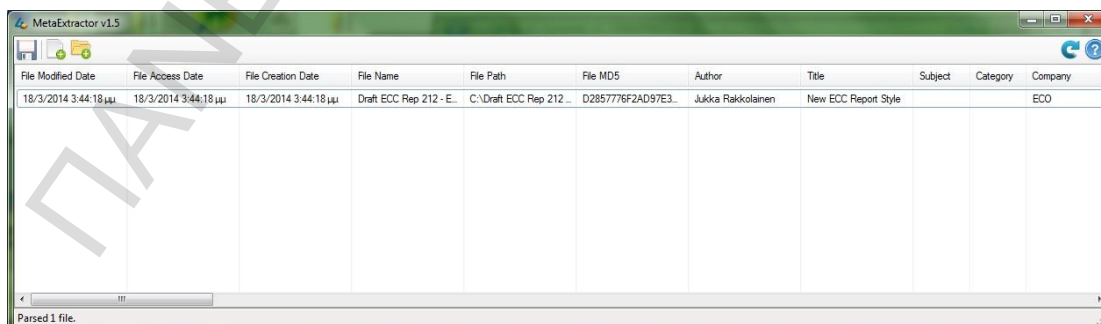
Θα διαπιστώσουμε ότι ακόμα και χωρίς το άνοιγμα του αρχείου, συνεπώς και μέσω διαφορετικού λειτουργικού συστήματος ή και με απουσία του λογισμικού με το οποίο δημιουργήθηκε, μπορούμε να έχουμε αρκετές πληροφορίες, που πιθανόν να χρησιμεύουν σε κάποιον που εξετάζει ένα αρχείο ή σε κάποιον που επιθυμεί να προστατεύσει τα προσωπικά του δεδομένα από κακόβουλη χρήση.

11.1 Αυτοματοποιημένα με χρήση εργαλείου ελεύθερης χρήσης

Σε αυτή την ενότητα θα χρησιμοποιήσουμε ενδεικτικά δύο από τα εργαλεία, που διανέμονται στο διαδίκτυο ελεύθερα προς χρήση, και που αναφέραμε στο προηγούμενο κεφάλαιο, και θα προσπαθήσουμε να εξάγουμε μεταδεδομένα των δύο τύπων εγγράφων.

11.1.1 Μεταδεδομένα αρχείου .docx με το εργαλείο MetaExtractor

Αφού επιλέξουμε το αρχείο από το παράθυρο διαλόγου που εμφανίζεται κατά το άνοιγμα, η ανάλυση εκτελείται αυτόματα και παρουσιάζονται στο παράθυρο τα αποτελέσματα, που όμως δεν είναι ευανάγνωστα χωρίς να κάνουμε οριζόντια κύλιση της μπάρας.

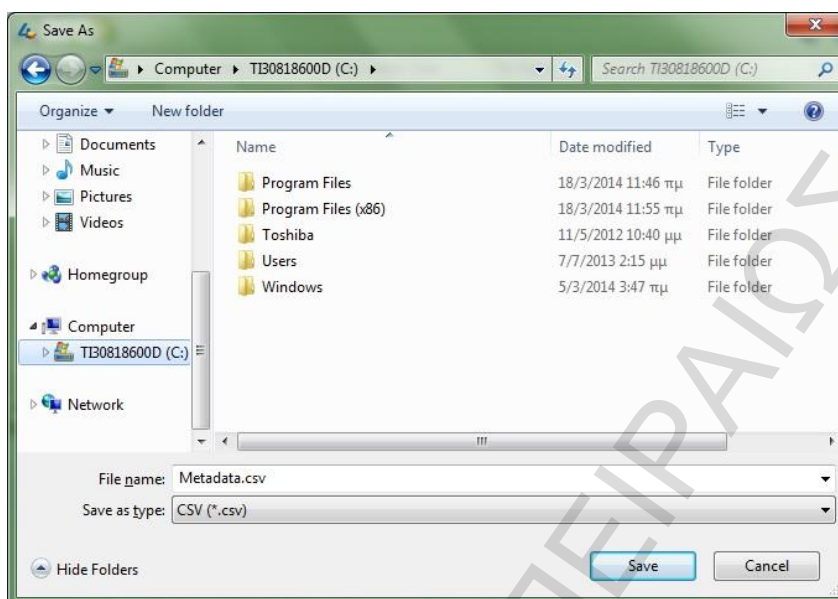


File Modified Date	File Access Date	File Creation Date	File Name	File Path	File MD5	Author	Title	Subject	Category	Company
18/3/2014 3:44:18 μμ	18/3/2014 3:44:18 μμ	18/3/2014 3:44:18 μμ	Draft ECC Rep 212 - E...	C:\Draft ECC Rep 212 ..	D2857776F2AD97E3...	Jukka Rakkolainen	New ECC Report Style			ECO

Εικόνα 25 Μεταδεδομένα που εντόπισε το MetaExtractor

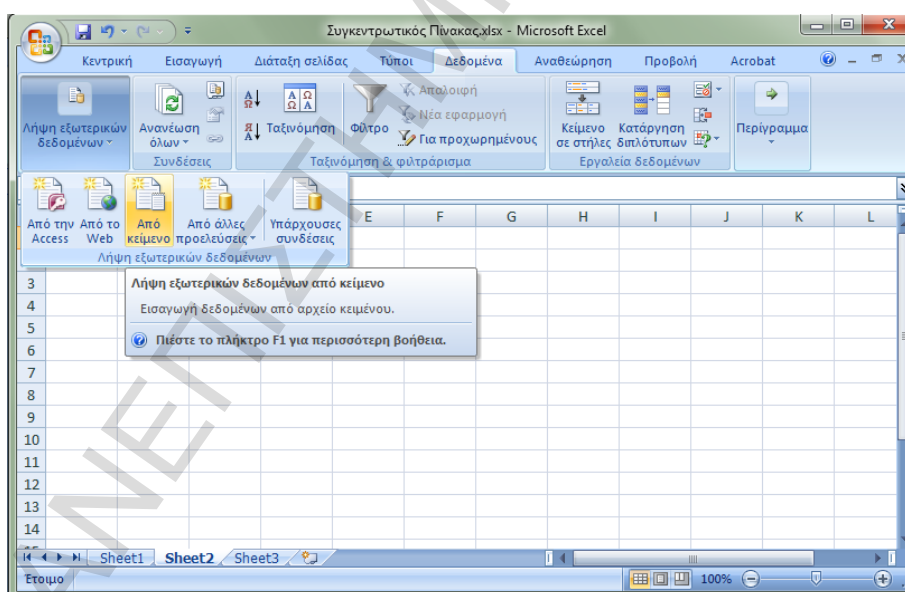
Πατώντας το κουμπί με τη μορφή δισκέτας (Save as), διαλέγουμε τη διαδρομή για την αποθήκευση των αποτελεσμάτων ως Τιμές διαχωριζόμενες με κόμμα (Comma

Separated Values *.csv) τις οποίες μπορούμε να εισάγουμε ως δεδομένα από κείμενο (IData from Text) στο Excel.



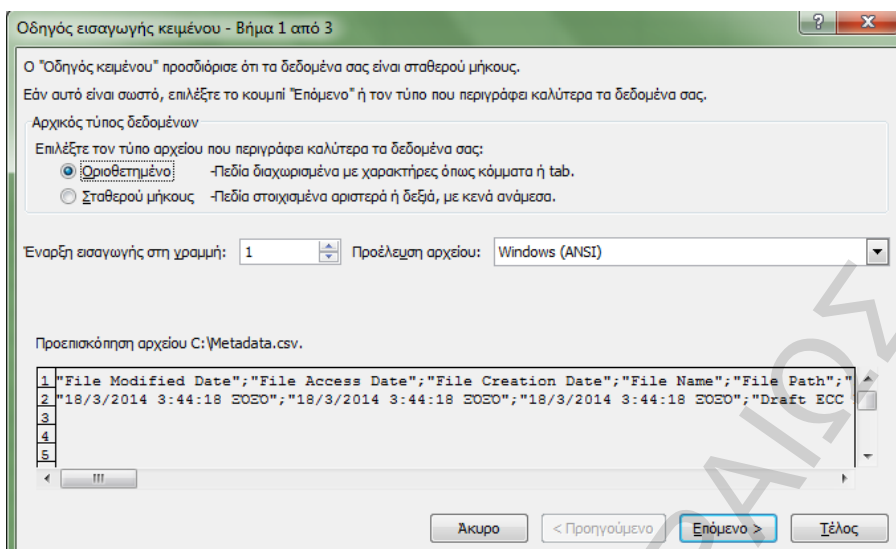
Εικόνα 26 Εξαγωγή και αποθήκευση των Μεταδεδομένων από το MetaExtractor

Ανοίγουμε το MS Excel 2007 και από την καρτέλα "Δεδομένα" επιλέγουμε "Λήψη εξωτερικών δεδομένων" και στη συνέχεια "Από κείμενο":



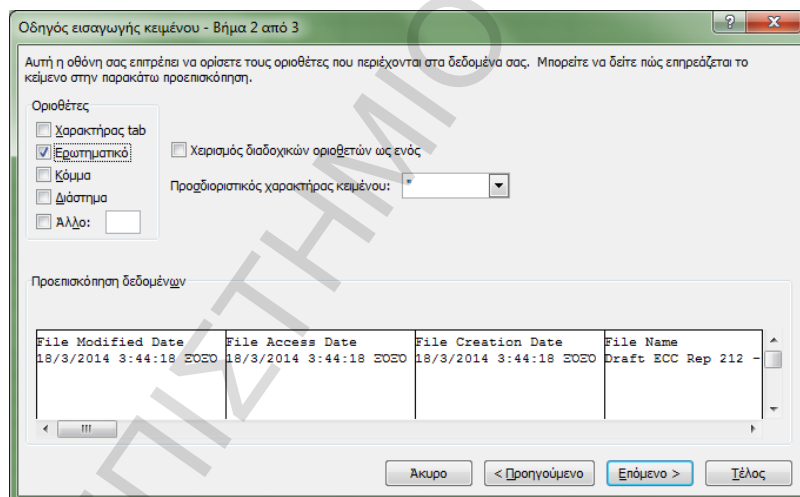
Εικόνα 27 Εισαγωγή μεταδεδομένων από το MetaExtractor στο MS Excel 2007

Έπειτα εμφανίζεται ο Οδηγός εισαγωγής κειμένου που μας βοηθά και καθοδηγεί για την ενσωμάτωση των τιμών, που έχουμε αποθηκεύσει ως CSV αρχείο, σε ξεχωριστά κελιά του Excel.



Εικόνα 28 Επιλογή του τύπου δεδομένων στο αρχείο που θέλουμε να εισάγουμε

Προσδιορίζουμε τα δεδομένα του αρχείου κειμένου που εισάγουμε ως "οριοθετημένο" με οριοθέτη το "ερωτηματικό" και το επιβεβαιώνουμε βλέποντας πώς παρουσιάζονται τα δεδομένα στο πλαίσιο προεπισκόπησης.



Εικόνα 29 Ορισμός οριοθέτη διαφορετικών τιμών εισαχθέντων δεδομένων

Στη συνέχεια, αφού εισαχθούν οι τιμές στα κελιά του Excel, αντιμετωπίζουμε τις γραμμές με τις στήλες, ώστε είναι πιο ευανάγνωστα.

Παρατηρούμε ότι παρέχονται τα παρακάτω δεδομένα:

- Ημερομηνία Τροποποίησης Αρχείου (File Modified Date)
- Ημερομηνία Προσπέλασης Αρχείου (File Access Date)
- Ημερομηνία Δημιουργίας Αρχείου (File Creation Date)
- Όνομα Αρχείου (File Name)
- Διαδρομή Αρχείου (File Path)
- (Αλγόριθμος) MD5 Αρχείου (File MD5)

- Συντάκτης (Author)
- Τίτλος (Title)
- Θέμα (Subject) [κενό πεδίο]
- Κατηγορία (Category) [κενό πεδίο]
- Εταιρία (Company)
- Λέξεις-κλειδιά (Keywords) [κενό πεδίο]
- Σχόλια (Comments) [κενό πεδίο]
- Κατάσταση (Status) [κενό πεδίο]
- Δημιουργήθηκε (Created)
- Τροποποιήθηκε (Modified)
- Τελευταία Τροποποίηση Από (Last Modified By)
- Αναθεώρηση (Revision)
- Ανενεργοί Σύνδεσμοι (Links Dirty) [κενό πεδίο]
- Διαχειριστής (Manager) [κενό πεδίο]
- Κλίμακα (Scale)
- Αριθμός Κλιπ Πολυμέσων (Multimedia Clip Count) [κενό πεδίο]
- Αριθμός Κρυφών Διαφανειών (Hidden Slide Count) [κενό πεδίο]
- Αριθμός Σημειώσεων (Note Count) [κενό πεδίο]
- Αριθμός Διαφανειών (Slide Count) [κενό πεδίο]
- Αριθμός Παραγράφων (Paragraph Count)
- Line Count (Αριθμός Γραμμών)
- Byte Count (Αριθμός Byte) [κενό πεδίο]
- Μορφή Παρουσίασης (Presentation Format) [κενό πεδίο]
- Κωδικός Σελίδας (Code Page) [κενό πεδίο]
- Ασφάλεια (Security)
- Όνομα Εφαρμογής (Application Name)
- Αριθμός Χαρακτήρων (Char Count)
- Αριθμός Λέξεων (Word Count)
- Αριθμός Λέξεων (Page Count)
- Τελευταία Εκτύπωση (Last Printed) [κενό πεδίο]
- Τελευταία Εκτύπωση από (Last Printed By) [κενό πεδίο]
- Χρόνος Επεξεργασίας (Edit Time)
- Πρότυπο (Template)
- Όνομα Αρχείου Προτύπου (Template Filename) [κενό πεδίο]
- Αριθμός Χαρακτήρων με Διαστήματα (Char With Spaces Count)
- Διαμοιρασμένο Έγγραφο (Shared Doc)
- Έκδοση Εφαρμογής (Application Version)
- Σύνδεσμοι Έως Σήμερα (Links Up To Date)
- Αλλαγμένοι Υπερσύνδεσμοι (Hyperlinks Changed)
- Εφαρμογή Παραγωγού (Producer Application) [κενό πεδίο]
- Τελευταίο Όνομα Αποθήκευσης (Last Save Name) [κενό πεδίο]

- Τελευταία Διαδρομή Αποθήκευσης (Last Save Path) [κενό πεδίο]
- Αριθμός Αντικειμένων (Object Count) [κενό πεδίο]
- Αριθμός Αντικειμένων OLE (Ole Object Count) [κενό πεδίο]
- Αριθμός Πινάκων (Table Count) [κενό πεδίο]
- Αριθμός Κελιών (Cell Count) [κενό πεδίο]
- Αριθμός Εικόνων (Image Count) [κενό πεδίο]
- Αριθμός Σχημάτων (Draw Count) [κενό πεδίο]
- Αριθμός Πλαισίων (Frame Count) [κενό πεδίο]
- Γλώσσα (Language) [κενό πεδίο]
- Προσαρμοσμένα Δεδομένα (Custom Data) [κενό πεδίο]

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ

File Modified Date	18/3/2014 3:44:18 ΞΟΞΟ
File Access Date	18/3/2014 3:44:18 ΞΟΞΟ
File Creation Date	18/3/2014 3:44:18 ΞΟΞΟ
File Name	Draft ECC Rep 212 - Evolution in the use of E 212 MNCs - with amended annex.docx
File Path	C:\Draft ECC Rep 212 - Evolution in the use of E 212 MNCs - with amended annex.docx
File MD5	D2857776F2AD97E30270D9C96895CE25
Author	Jukka Rakkolainen
Title	New ECC Report Style
Subject	
Category	
Company	ECO
Keywords	
Comments	
Status	
Created	11/11/2013 12:36:00 ΞΟΞΟ
Modified	11/11/2013 12:36:00 ΞΟΞΟ
Last Modified By	Freddie McBride
Revision	2
Links Dirty	
Manager	
Scale	FALSE
Multimedia Clip Count	
Hidden Slide Count	
Note Count	
Slide Count	
Paragraph Count	143
Line Count	515
Byte Count	
Presentation Format	
Code Page	
Security	0
Application Name	Microsoft Office Word
Char Count	61850
Word Count	10139
Page Count	30
Last Printed	
Last Printed By	
Edit Time	0:01:00
Template	Normal
Template Filename	
Char With Spaces Count	71846
Shared Doc	FALSE
Application Version	140.000
Links Up To Date	FALSE
Hyperlinks Changed	FALSE
Producer Application	
Last Save Name	
Last Save Path	
Object Count	
Ole Object Count	
Table Count	
Cell Count	
Image Count	
Draw Count	
Frame Count	
Language	
Custom Data	

Εικόνα 30 Παρουσίαση των μεταδεδομένων από το MetaExtractor στο Excel

11.2 Metadata Touch

The screenshot displays the Metadata Touch interface with the following fields and values:

Keep	Title:	Office Mda
Keep	Subject:	Office medatada
Change	Author:	Πάσχος Χριστόδουλος
Append	Company:	UniPei
Prepend	Manager:	
Remove	Keywords:	
Keep	Category:	
Keep	Status:	
Keep	Comments:	
Keep	Last modified by:	Chris
Keep	Revision number:	2
Keep	Total editing time:	1 Minutes
Keep	Template:	Normal.dotm
Keep	Application:	Microsoft Office Word
Keep	Application version:	14.000

Εικόνα 31 Παρουσίαση των μεταδεδομένων από το Metadata Touch

Όπως φαίνεται και στην παραπάνω εικόνα το interface της εφαρμογής είναι ιδιαίτερα χρηστικό. Παρουσιάζονται όλα τα πεδία των μεταδεδομένων που έχουμε αναλύσει και δίνεται η δυνατότητα να τα αφαιρέσουμε, να τα τροποποιήσουμε κτλ.

12 Σύνοψη

Τα μεταδεδομένα που περιέχονται σε έγγραφα, έχουν πράγματι χρησιμότητα σε πολλές περιπτώσεις, όπως η διευκόλυνση της συνεργασίας μεταξύ μιας ομάδας ανθρώπων. Παρόλα αυτά πολλοί έχουν άγνοια για τον τύπο πληροφορίας που αποθηκεύεται στα έγγραφα, φύλλα εργασίας ή παρουσιάσεις τους. Μεγάλη προσοχή απαιτείται από τους υπεύθυνους χρήστες για την διαφύλαξη της μη διαρροής ευαίσθητων πληροφοριών σε τρίτους.

Σε αυτή την εργασία, παρουσιάσαμε αναλυτικά μια πληθώρα μεταδεδομένων που αποθηκεύονται σε αρχεία του Microsoft Office, είδαμε ενδεικτικά πώς μπορούμε να τα εξάγουμε ή να τα εντοπίσουμε και τρόπους για να τα απομακρύνουμε.

Η κατανόηση των μεταδεδομένων και η υλοποίηση προληπτικών μέτρων ελέγχου είναι απαραίτητες για την προστασία των ευαίσθητων πληροφοριών. Μέχρι όμως οι χρήστες αποκτήσουν περισσότερη συνείδηση της κρυμμένης πληροφορίας που περιέχεται και διαμοιράζεται με τα έγγραφά τους, οι ερευνητές μπορούν και θα συνεχίσουν να εκμεταλλεύονται τα μεταδεδομένα των εγγράφων κατά τη διενέργεια μιας διερεύνησης.

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΡΠΕΡ

Βιβλιογραφία

Διαδικτυακοί Τόποι

- [1] <http://office.microsoft.com/>
- [2] <http://support.microsoft.com/>
- [3] <http://msdn.microsoft.com/>
- [4] <http://computer-forensics.sans.org/blog/2009/07/10/office-2007-metadata/>
- [5] <http://www.symantec.com/connect/articles/microsoft-office-security-part-two>
- [6] <http://introductiontocomputerforensics.blogspot.gr/2012/05/office-x-and-forensics.html>
- [7] http://blogs.technet.com/b/office_resource_kit/archive/2009/07/07/what-document-inspector-doesn-t-catch.aspx

Άρθρα

- [8] Document Metadata and Computer Forensics, Jeffrey R. Jones (James Madison University Infosec Techreport)
- [9] Forensic Implications of Metadata in Electronic Files, by John Ruhnka (University of Colorado) and John W. Bagby (Pennsylvania State University)
- [10] Simson L. Garfinkel, James J. Migletz: New XML-Based Files Implications for Forensics. IEEE Security & Privacy 7(2): 38-44 (2009)

Εργασίες

- [11] Automated Metadata Extraction, James Migletz (Naval Postgraduate School Monterey, California), June 2008