



PDF:

οδηγίες για τη μετατροπή από αρχείο
κειμένου docx μέσω της εφαρμογής του
Word (MSOffice)

2.1 έκδοση, 2022

Σειρά εγχειριδίων: Παραγωγή προσβάσιμων ηλεκτρονικών μορφότυπων
για την ψηφιακή βιβλιοθήκη της AMELib

Στοιχεία έκδοσης

Τίτλος: PDF: οδηγίες για τη μετατροπή από αρχείο κειμένου docx μέσω της εφαρμογής του Word (MSOffice)

Έκδοση: 2.1

Έτος: Σεπτέμβριος 2022

Σειρά: Παραγωγή προσβάσιμων ηλεκτρονικών μορφότυπων για την ψηφιακή βιβλιοθήκη της AMELib

Σύνταξη & επιμέλεια: Ιωάννα Ζορμπά (Βιβλιοθήκη & Κέντρο πληροφόρησης του Πανεπιστημίου Θεσσαλίας)

copyright © Σ.Ε.Α.Β. (AMELib)



Αυτή η εργασία χορηγείται με άδεια Creative Commons

[Αναφορά Δημιουργού-Μη Εμπορική Χρήση-Όχι Παράγωγα Έργα 4.0 Διεθνές](https://creativecommons.org/licenses/by-nc-nd/4.0/)

URL: <http://creativecommons.org/licenses/by-nc-nd/4.0/>

Περιεχόμενα

Στοιχεία έκδοσης	2
Περιεχόμενα	3
Τι είναι ένα .pdf αρχείο και κατά πόσο προσβάσιμο μπορεί να γίνει; ...	4
Προσβασιμότητα αρχείων pdf.....	5
Προτάσεις βελτίωσης προσβασιμότητας	7
Προσθήκη μεταδεδομένων	8
Επεξεργασία στυλ (DAISY)	8
Προσθήκη δυναμικού πίνακα περιεχομένων	9
Μέγεθος & απόσταση γραμματοσειράς	9
Διάστημα παραγράφων	10
Πίνακες	10
Υποσημειώσεις	11
Εξαγωγή σε pdf αρχείο από MSWord 2016 και μετέπειτα εκδόσεις	12
Έλεγχος προσβασιμότητας μέσω Adobe Professional	14
Ανάγνωση αρχείων pdf	14

Τι είναι ένα .pdf αρχείο και κατά πόσο προσβάσιμο μπορεί να γίνει;

Ένα pdf (Portable Document Format) αρχείο (.pdf) είναι ένα αρχείο μορφοποιημένου κειμένου, που μπορεί να φιλοξενήσει εικόνες και υπερσυνδέσμους, αλλά και άλλα πολυμεσικά αντικείμενα που διαθέτουν ήχο και βίντεο. Έχει καθιερωθεί ως επίσημο πρότυπο για σκοπούς εμπορικούς ή διοικητικούς του δημοσίου με δυνατότητες ασφαλούς διακίνησης της πληροφορίας και της μορφής της. Είναι ένας πλέον διαδεδομένος τύπος αρχείων διάχυσης πληροφορίας και διατίθεται πληθώρα λογισμικών ανάγνωσης (readers) για όλες τις συσκευές και τα λειτουργικά συστήματα. Χρησιμοποιείται για ηλεκτρονικές εκδόσεις, αλλά ο σχεδιασμός του έχει γίνει ώστε να λειτουργεί ως αρχείο έτοιμο για εκτύπωση. Γι' αυτό και οι διαστάσεις των σελίδων καθορίζονται από το μέγεθος του χαρτιού, όπως το A4.

Κάποια από τα χαρακτηριστικά του είναι:

- Αποτελείται από ένα και μόνο αρχείο και όχι από σύνολο αρχείων σε συμπιεσμένη μορφή (.zip/.rar), όπως είναι τα daisy books ή οι web εκδόσεις.
- Μπορεί να καταχωρήσει ορισμένα μεταδεδομένα, όπως τίτλο, συντάκτη, θέμα και δικαιώματα.
- Μπορεί να έχει μικρό μέγεθος, ανάλογα με τον τρόπο παραγωγής και επεξεργασίας του.
- Η πλοήγησή του είναι και σειραϊκή, αλλά μπορεί να επιτυγχάνεται εύχρηστα με τη χρήση σελιδοδεικτών (bookmarks) για στοιχειοθέτηση λίστας περιεχομένων. Επίσης, η ύπαρξη tags για τις επικεφαλίδες <H> του

κειμένου συμβάλλουν στην προσβασιμότητα μέσω υποστηρικτικών λογισμικών. Τέλος, επιτυγχάνεται αναζήτηση μέσα στο περιεχόμενο από την στιγμή που εφαρμόζεται οπτική αναγνώριση χαρακτήρων (OCR).

Σημειώνεται ότι ένα pdf αρχείο μπορεί να είναι παράγωγο σάρωσης (scan) ή διαδικασίας εκτύπωσης (print) και να έχει ουσιαστικά μόνο την ιδιότητα της εικόνας (image). Σε αυτήν την περίπτωση το κείμενο δεν αναγνωρίζεται (επιλέγεται) και δεν μπορεί να αναζητηθεί ή να αποδοθεί ακουστικά με συνθετική φωνή. Χρειάζεται περαιτέρω επεξεργασία του, για να μετατραπεί σε αναγνώσιμο, άρα και πιο προσβάσιμο. Τέλος, συνήθως, το μέγεθος του αρχείου είναι συγκριτικά μεγαλύτερο σε σχέση με ένα pdf εξαγόμενο από αρχείο κειμένου.

Προσβασιμότητα αρχείων pdf

Ένα pdf αρχείο μπορεί να θεωρηθεί ότι έχει στοιχεία προσβασιμότητας όταν διαθέτει:

- Αναγνώριση χαρακτήρων (OCR - Optical Character Recognition)
- Συνεχές κείμενο, απαλλαγμένο από στήλες και διπλές (αντικριστές) σελίδες
- Διαρθρωμένο περιεχόμενο μέσω bookmarks
- Δομημένο περιεχόμενο με tags επικεφαλίδων, εικόνων και πινάκων με εναλλακτικό κείμενο (βλέπε σχετικά, <https://youtu.be/m1XMJZEnVZY> (Tagging a PDF using Adobe Acrobat DC))

Σημειώνεται ότι τα pdf αρχεία δεν θεωρούνται από τους πιο προσβάσιμους μορφότυπους, γιατί παρουσιάζουν τους παρακάτω τουλάχιστον περιορισμούς:

- Το κείμενό τους δεν μπορεί να αναπροσαρμόζεται ανάλογα με το μέγεθος του πλαισίου εμφάνισής του ή ανάλογα με προσωποποιημένες ρυθμίσεις εμφάνισης γραμματοσειράς, διαστημάτων μεταξύ γραμμάτων και λέξεων ή γραμμών και παραγράφων. Οι αναγνώστες των pdf αρχείων μπορεί να βελτιώνονται σε θέματα προσβασιμότητας, αλλά υπάρχουν ανυπέρβλητοι περιορισμοί στον μορφότυπο αυτόν. Συνεπώς, χρειάζεται το πρωτογενές κείμενο από το οποίο θα εξαχθεί το pdf να είναι διαμορφωμένο έτσι, ώστε να είναι φιλικό/προσβάσιμο προς κάποιες κατηγορίες εντυποανάπηρων αναγνωστών (βλέπε: [προτάσεις βελτίωσης προσβασιμότητας](#)).
- Οι παραπομπές σε υποσημειώσεις και σημειώσεις τέλους παρουσιάζουν τους παρακάτω περιορισμούς:
 - Δεν είναι δυναμικές, άρα δεν δίνεται δυνατότητα μετάβασης από/προς αυτές.
 - Δεν αναγνωρίζονται από τους αναγνώστες οθόνης. Συνεπώς, η πληροφορία ύπαρξης παραπομπής παρακάμπτεται τελείως, όταν ο αναγνώστης ακούει το αρχείο.
 - Το κείμενο των υποσημειώσεων ακούγεται συνεχόμενα μετά το τέλος του κυρίως κειμένου της σελίδας, δημιουργώντας σύγχυση στον αναγνώστη που ακούει το κείμενο. Επίσης, το πλήρες περιεχόμενο μιας υποσημείωσης μπορεί να ολοκληρώνεται στο υποσέλιδο της επόμενης

σελίδας. Αυτός ο διαμερισμός και τεμαχισμός της πληροφορίας μπορεί να δυσκολέψει και τους αναγνώστες με δυσκολία χειρισμού άνω άκρων ή με μερική όραση, αφού χρειάζεται να προβαίνουν σε επιπλέον χειρισμούς για τον εντοπισμό της συνέχειας της πληροφορίας μέσα σε μία ή και περισσότερες σελίδες (βλέπε: [προτάσεις βελτίωσης προσβασιμότητας](#)).

- Η ηχητική απόδοση με συνθετική φωνή στα ελληνικά κείμενα για τους περισσότερους αναγνώστες pdf προϋποθέτει παράλληλη λειτουργία λογισμικού αναγνώστη οθόνης (screen reader), όπως είναι για παράδειγμα το NVDA (<https://www.nvaccess.org>) ή τον αναγνώστη οθόνης των Windows «Narrator».

Ενδεικτικές πηγές από την Adobe για την προσβασιμότητα των pdf αρχείων:

- <https://www.adobe.com/accessibility.html> (Accessibility)
- <https://www.adobe.com/accessibility/pdf/pdf-accessibility-overview.html> (PDF Accessibility Overview)

Προτάσεις βελτίωσης προσβασιμότητας

Προτού προχωρήσουμε σε [εξαγωγή σε pdf από MSWord 2016 και μετά](#), μπορούμε να προβούμε σε μερικές καθοριστικές βελτιώσεις προσβασιμότητας, ώστε να μειώσουμε όσο περισσότερους περιορισμούς ή δυσκολίες μπορούμε προς όφελος εντυποανάπηρων αναγνωστών.

Σε ένα **αντίγραφο** του τελικού επεξεργασμένου κειμένου μορφής docx μπορούμε να εφαρμόσουμε όλες ή ορισμένες από τις παρακάτω πρακτικές και από αυτό να εξάγουμε το προσβάσιμο pdf αρχείο.

Ειδικότερα για ζητήματα δυσλεξίας, οι παρακάτω προτάσεις έχουν στηριχτεί στο: British Dyslexia Association. *Dyslexia Style Guide 2018: Creating Dyslexia Friendly Content*.

Accessed August 1, 2021.

<https://www.bdadyslexia.org.uk/advice/employers/creating-a-dyslexia-friendly-workplace/dyslexia-friendly-style-guide>

Προσθήκη μεταδεδομένων

Μέσω επεξεργασίας Word δεν μπορούμε να προσθέσουμε πλήρη μεταδεδομένα. Ωστόσο, χρειάζεται να συμπληρώσουμε -αν δεν το έχουμε ήδη κάνει κατά την επεξεργασία του παραδοτέου docx- για λόγους προσβασιμότητας τουλάχιστον ορισμένα που διατίθενται από τις **Ιδιότητες**, όπως Τίτλος, Συντάκτης και Θέμα. Όταν έχουμε πάνω από έναν συγγραφέα ή θέμα, διαχωρίζουμε τα δεδομένα με το σύμβολο του ελληνικού ερωτηματικού (semicolon) με κενό ή απλό κόμμα με κενό.

Επεξεργασία στυλ (DAISY)

Απαλείφουμε τα στυλ DAISY από στοιχεία μορφοποίησης χρωματικού φόντου και γραμματοσειράς, καθώς και περιγράμματος. Οι αναγνώστες pdf μπορεί να δίνουν δυνατότητα εμφάνισης του κειμένου με συνδυασμό χρώμα φόντου και γραμματοσειράς. Συνεπώς, απαιτείται επεξεργασία των στυλ.

Προσθήκη δυναμικού πίνακα περιεχομένων

Εισαγάγουμε μετά τη σελίδα τίτλου πίνακα περιεχομένων με βάση τις Επικεφαλίδες του αρχείου. Προσέχουμε τον τίτλο της ενότητας σε σχέση με εκείνη των περιεχομένων της έντυπης έκδοσης. Προτείνεται για λόγους τεκμηρίωσης να μην προχωρούμε σε διαγραφή των σελίδων με τα περιεχόμενα της έντυπης έκδοσης.

Μέγεθος & απόσταση γραμματοσειράς

Οι προδιαγραφές για την επεξεργασία κειμένου έχουν ήδη καλύψει καλές πρακτικές προς όφελος των αναγνωστών με δυσλεξία σε ό,τι αφορά:

- Χρήση τύπου γραμματοσειράς sans serif fonts, όπως το arial
- Μέγεθος γραμματοσειράς μεταξύ 12-14
- Στοίχιση κειμένου αριστερά
- Μονόστηλο κείμενο
- Διάστιχο 1,5
- Αποφυγή πλάγιας γραφής
- Αποφυγή πλεονασματικής χρήσης έντονης γραφής
- Αποφυγή χρήσης κεφαλαίων γραμμάτων και μικρών κεφαλαίων (small caps)

Ωστόσο, μπορούμε να βελτιώσουμε το μέγεθος της γραμματοσειράς για το βασικό κείμενο και τις επικεφαλίδες.

Προτείνεται:

- Βασικό κείμενο 13 μέγεθος γραμματοσειράς
- Μέγεθος γραμματοσειράς για επικεφαλίδες 15, ώστε να πετύχουμε την καλή πρακτική το μέγεθος των

επικεφαλίδων να είναι 20% μεγαλύτερο από το βασικό κείμενο

- Απόσταση γραμματοσειράς εκτεταμένη κατά 1 στ.

Έτσι συνεπαγωγικά πετυχαίνουμε το εύρος των χαρακτήρων σε μία γραμμή να είναι μέχρι 60 χαρακτήρες, ικανοποιώντας ακόμη μία προδιαγραφή.

Διάστημα παραγράφων

Για να βελτιώσουμε την κατανόηση του κειμένου από αναγνώστες με δυσλεξία ή με μειωμένη όραση, μπορούμε να εφαρμόσουμε την προτεινόμενη πρακτική για διαστήματα πριν και μετά τις παραγράφους και τις επικεφαλίδες.

Συγκεκριμένα:

- Για το βασικό κείμενο, διάστημα πριν και μετά 6 ή 12 στ.
- Για επικεφαλίδες, διάστημα πριν 12 στ. και μετά 6 ή 12 στ.

Πίνακες

Ελέγχουμε τους πίνακες του κειμένου και εφαρμόζουμε:

- Έντονη γραφή στη γραμμή κεφαλίδας, όταν δεν εφαρμόζεται
- Αλλαγή προσανατολισμού σελίδας για την περίπτωση πίνακα με πολλές στήλες ή με εκτενές κείμενο εντός των κελιών, ώστε να τον κάνουμε πιο ευανάγνωστο αξιοποιώντας το εύρος της σελίδας
- Στην περίπτωση που παρουσιάζεται αναδίπλωση λέξεων εντός των κελιών, τότε κάνουμε **επιλεκτικά** χρήση μη αυτόματου συλλαβισμού από την καρτέλα Διάταξη

- Επιλογή επανάληψης γραμμής κεφαλίδας
- Απενεργοποίηση επιλογής επέκτασης γραμμής σε δύο σελίδες

Υποσημειώσεις

Για να αποφύγουμε τις δυσλειτουργίες σχετικά με τις υποσημειώσεις και σημειώσεις τέλους, που αναφέρθηκαν πιο πάνω στα θέματα [προσβασιμότητας](#) των pdf, προτείνονται δύο τροποποιήσεις:

α) Μεταφορά όλων των υποσημειώσεων ή σημειώσεων τέλους σε ξεχωριστή ενότητα στο τέλος του κειμένου με ενδεικτικό τίτλο «Υποσημειώσεις κυρίως κειμένου» σε στυλ Επικεφαλίδα 1. Κάθε υποσημείωση αριθμείται εκ νέου με αύξουσα αριθμητική σειρά και αποκτά τίτλο «Υποσημείωση x» σε στυλ Επικεφαλίδα 2. Έτσι πετυχαίνουμε την πλοήγηση σε αυτές. Η παραπομπή εντός του κυρίως κειμένου με τη χρήση εκθέτη αντικαθίσταται από τυποποιημένη φράση {Υποσημείωση x}. Στις γενικές μετατροπές επί του κειμένου, που ορίζονται στην αρχή, προσθέτουμε και την παρακάτω σύμβαση:

«Οι υποσημειώσεις έχουν μεταφερθεί σε ξεχωριστή ενότητα στο τέλος του κειμένου και η κωδικοποίησή τους έχει μετατραπεί σε συνεχής αριθμητική. Εντός κειμένου υπάρχει τυποποιημένη παραπομπή σε καθεμία από αυτές προς αντικατάσταση του εκθέτη. Δεν δίνεται η δυνατότητα αυτόματης μετάβασης από/προς αυτές.»

β) Μετατροπή των υποσημειώσεων σε σημειώσεις τέλους με την προσθήκη της παρακάτω σύμβασης:

«Οι υποσημειώσεις έχουν μεταφερθεί στο τέλος του κειμένου ως σημειώσεις τέλους και η κωδικοποίησή τους έχει μετατραπεί σε συνεχής αριθμητική. Δεν δίνεται η δυνατότητα αυτόματης μετάβασης από/προς αυτές.»

Εξαγωγή σε pdf αρχείο από MSWord 2016 και μετέπειτα εκδόσεις

Ο στόχος μας είναι να διαθέτουμε προσβάσιμους μορφότυπους. Συνήθως τα pdf που χρησιμοποιούμε ως πρότυπο/βάση για τη μετατροπή του περιεχομένου σε άλλους μορφότυπους δεν έχουν καθόλου ή δεν έχουν πλήρη περιεχόμενα (βλέπε, bookmarks & heading tags). Επίσης, δεν συμπεριλαμβάνουν εναλλακτικό κείμενο σε εικόνες και πίνακες. Η επεξεργασία τους με εμπορικά λογισμικά είναι δαπανηρή και στην απόκτησή τους και στον απαιτούμενο χρόνο επεξεργασίας. Συνεπώς, προτείνεται η εξαγωγή pdf αρχείου από το επεξεργάσιμο docx σύμφωνα με τις οδηγίες του Δικτύου της AMELib. Ένα αρχείο επεξεργασμένο σύμφωνα με αυτές τις οδηγίες μπορεί να παράγει ένα αξιόπιστα προσβάσιμο pdf αρχείο και, ακολουθώντας τις προτάσεις βελτίωσης, μπορεί να παράγει σχεδόν 100% προσβάσιμο pdf. Αν διαθέτουμε το [Adobe Acrobat Professional](#) μπορούμε να προβούμε σε έλεγχο προσβασιμότητας και ίσως σε μικρές βελτιώσεις.

Τα **βήματα εξαγωγής** του pdf από MSWord 2016 και μετέπειτα εκδόσεις είναι τα εξής:

Μετά την ολοκλήρωση της επεξεργασίας του αρχείου κειμένου σε docx και τη δημιουργία του αρχείου epub ή και του daisy book (για να αποκλείσουμε περαιτέρω/διορθωτική

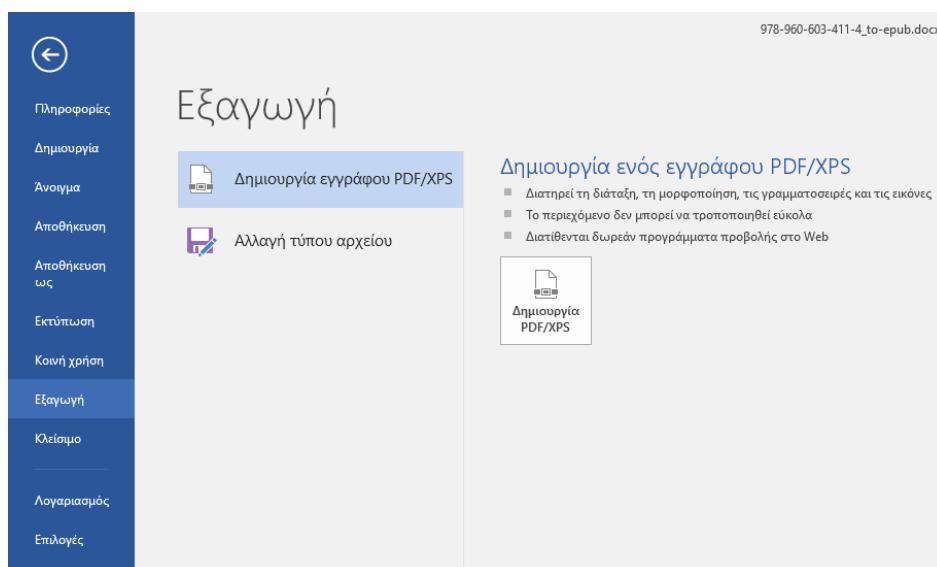
επεξεργασία του κειμένου), πραγματοποιούμε τις παρακάτω ενέργειες:

- «**Export/Εξαγωγή**» >>> «**Δημιουργία/Create PDF/XPS**»
- Πριν την αποθήκευση, ανοίγουμε τις «**Options (Επιλογές)**»
 - Ενεργοποιούμε την επιλογή «**Document Structure tags for Accessibility (Ετικέτες δομές εγγράφου για άτομα με ειδικές ανάγκες)**».

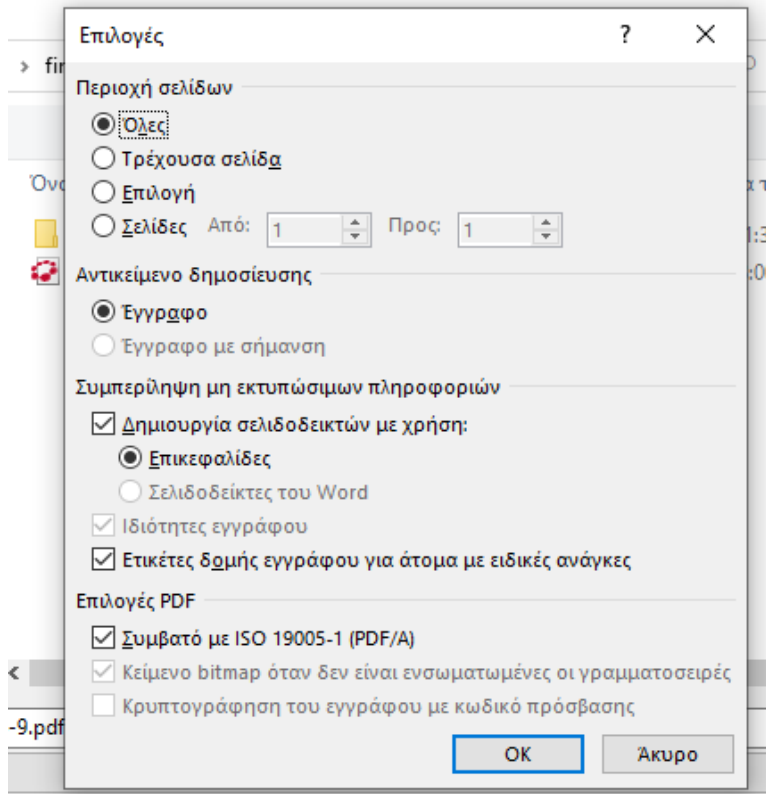
Προτείνεται το αρχείο να το ονομάζουμε ως ISBN_from_doc (παράδειγμα: 978-960-9552-29-5-1_from_doc.pdf), ώστε να διαφοροποιείται από το pdf-πηγή.

Βλέπε παρακάτω, σχετικές απεικονίσεις επιλογών και ρυθμίσεων.

Οθόνη επιλογής λειτουργίας «**Export/Εξαγωγή**» >> > «**Δημιουργία/Create PDF/XPS**»



Οθόνη επιλογής ρυθμίσεως εξαγωγής σε αρχείο pdf



Έλεγχος προσβασιμότητας μέσω Adobe Professional

Εάν διαθέτουμε έκδοση του εμπορικού λογισμικού Acrobat Professional της Adobe, μετά τον έλεγχο προσβασιμότητας προχωρούμε σε επιδιόρθωση τυχόν ζητημάτων/λαθών. Δεν αναμένονται σημαντικά λάθη, εφόσον ακολουθήθηκαν τα παραπάνω βήματα παραγωγής pdf αρχείου από επεξεργασμένο docx.

Ανάγνωση αρχείων pdf

Η ανάγνωση των αρχείων pdf μπορεί να επιτευχθεί από πολλούς αναγνώστες (readers) - ακόμη και από φυλλομετρητές διαδικτύου (web browsers) - των διαφορετικών συσκευών (desktop, tablet, smart phone), αλλά για να αξιοποιηθούν τα πλεονεκτήματα προσβασιμότητας,

όταν αυτά εφαρμόζονται, χρειάζεται επιλογή εκείνων των αναγνωστών που διαθέτουν τις αντίστοιχες δυνατότητες και φυσικά τις αντίστοιχες οδηγίες προς τους εν δυνάμει εντυποανάπηρους αναγνώστες (βλέπε για παράδειγμα: <https://helpx.adobe.com/acrobat/using/keyboard-shortcuts.html>).

Εκτός του Acrobat Reader, προτείνεται ο δωρεάν (free) αναγνώστης **Foxit** (<https://www.foxit.com/pdf-reader/>). Η παράλληλη λειτουργία του **NVDA** (www.nvaccess.org) αναγνώστη οθόνης είναι απαραίτητη.