

**תשובות לשאלות הבהרה פנייה לקבלת מידע מס 09/23 בדבר אספקת תוכנה לניהול ספריות בתי המשפט**

מס'	עמוד במכרז	סעיף במכרז	תוכן הבקשה להבהרה	תשובה
1.			<b>אפשרות חיפוש באמצעות OCR</b> – האם הכוונה למשתמשי הקצה (חיפוש באמצעות כלי גילוי) או במערכת הניהול עצמה?	האופטימאלי – להכניס את ה-OCR לשימוש פנימי וחיצוני (משתמשי קצה), על מנת שגם הספרנים יוכלו לבצע חיפושים טקסטואליים בחומרים סרוקים.
2.			מה ההבדל בין הסעיף האחרון (OCR) לבין סעיף 6 (אפשרות חיפוש במסמכים סרוקים)?	כעיקרון סעיף 11 נכלל בסעיף 6. כלומר, מערכת המשתמשת בטכנולוגיית ה-ORC לחיפוש במסמכים סרוקים ובתמונות – מהווה יתרון. עם זאת, כל מערכת לאחזור מידע רשאית לענות למכרז גם אם משתמשת בטכנולוגיה אחרת (לא מסוג OCR) ובתנאי שתיתן פתרון לחיפוש בחומר סרוק.
3.			האם סעיף שיתוף המידע עוסק גם במידע מודפס (השאלה בינספריתית בין היחידות השונות) או רק במידע אלקטרוני?	האופטימאלי – לאפשר לספרנים שקיפות מידע מלאה, הכוללת צפייה במשאבי הדפוס : מצאי הספריות השונות, מעבר עותקים בין הספריות (העברה בין-ספרייתית), כמו גם שקיפות מידע ממוחשב מכל סוג שהוא : תמונות/קבצים/סריקה/תמונות כריכה/קישור לספר/קישור למאגרים.

**כמו כן נבקש להבהיר מונחים מרכזיים במכרז :**

**הסבר מונחים קשורים למכרז :**

**Tagged d Image File Format** - (בראשי תיבות TIF או TIFF, לפעמים מתורגם כקובץ "טיף") הוא פורמט אחסון (container) לקובצי תמונות מפת סיביות.

שיטת המיפוי של TIFF מבוססת על פיקסלים והיא כוללת תמיכה בריבוי שכבות.

TIFF הוא פורמט גמיש יחסית, אשר יכול להכיל קבצים באיכות גבוהה, תומך בצבעים מ-1 ועד 24 ביט, תומך ברקע שקוף, שומר את המידע על התמונות, תומך בכמה אלגוריתמים של דחיסה ללא איבוד מידע וניתן לשימוש במכשירי ובתכנות פקס, סרוקים, מדפסות ותכנות לזיהוי כתב או לפרוק ועיבוד תמונה. לכן הוא פופולרי בקרב גרפיקאים, בדפוס, צלמים ובתעשיית ההוצאה לאור.

הפורמט נוצר ופותח במקור בשנות השמונים על ידי חברת אלדוס (Aldus) לשימוש בתוכנת הוצאה לאור שולחנית מתוצרתה וכיום הוא בבעלות חברת אדובי מערכות (Adobe Systems).

## זיהוי תווים אופטי - (Optical character recognition)

זיהוי תווים אופטי היא טכנולוגיה להמרת תמונה, טקסט מודפס וכתב יד שנסקרו על ידי סורק, למסמך תמליל ממוחשב.

תוכנות לזיהוי תווים אופטי, מתמודדות על פי רוב בהצלחה עם המרה של תמונה בה מצולם או סרוק מסמך מודפס אך רק לעיתים נדירות גם בכתב יד.

שימוש בטכנולוגיה של זיהוי תווים אופטי, והמרת קובץ גרפי טקסטואלי לקובץ המורכב מתווים, נעשים בדרך כלל לשם המטרות הבאות:

- עיבוד מחדש וסידור של התמליל, שילובו בתמליל מודפס עתידי, או הדפסתו מחדש בעיצוב אחר
- אחסון נתונים טקסטואליים במחשב, תוך הנגשת האפשרות של חיפוש תוכנם לחיפוש, למשל בתוכנת החיפוש הפנימית או במנועי חיפוש כדוגמת גוגל
- הפצת תוכן טקסטואלי עם תוכנה לאחזור מידע
- העלאת הטקסט לאתרי אינטרנט, או שליחתו בדואר אלקטרוני
- שינוי עיצוב התמליל לשם הפיכתו לנגיש עבור אנשים עם לקויות:
  - הדפסת התמליל או הצגתו על מרקע המחשב עבור אנשים עם לקות ראייה באותיות גדולות על דף צר, בצבעים בעלי ניגודיות חזקה (למשל צהוב על שחור), תוך הגדלת הרווחים בין המילים או בין השורות ביחס לגודל המילים, תוך ניקוי כתמים מהדף המקורי וכיוצא בזה.
  - הקראת התמליל על ידי המחשב בעזרת תוכנת הקראה (Text To Speech) עבור עיוורים ודיסלקטיים קשים
- הזנה למחשב של טופסי נייר באופן ממוכן

הערות:

- טעינת הקובץ למערך לכאורה היא מיותרת, ומבזבזת זמן ומקום בזיכרון המחשב. אבל תהליך החיפוש, מהיר במערך מאשר בקובץ.

**תוכנות של ocr : hocr, תוכנה חופשית לפיענוח אופטי עברי (ראו hOCR, שהוא שמו של תקן סימון מאוחר יותר למידע טקסטואלי שחולץ בתהליך זיהוי אופטי)**

**tesseract-ocr, תוכנה בקוד פתוח ל-OCR שפותחה על ידי HP וכעת על ידי גוגל OcrOpus, תוכנה בקוד פתוח ל-OCR**