

תקציר

למרות תחזית ארוכת הטווח שמחשבים יחליפו בצורה מוחלטת כתב יד בכל צורותיו, השימוש בדף ועט עודנם נפוצים להעברת אינפורמציה בחיי היום-יום. המרת כתב היד לצורה דיגיטלית טומנת בחובה יתרונות רבים. טקסט ממחושב לא רק מאפשר שיכפול יעיל אלא יכולות עיבוד נרחבות כגון, עריכה חיפוש, מיון, שיתוף וכו'.

בעית זיהוי כתב-יד בשפה הערבית בצורה יעילה נחשב לבעיה מאתגרת וזאת מאחר והשפה הערבית נכתבת בצורה מחוברת. לאחר תקופה ממושכת של עשרות שנים בה חוקרים התעניינו בזיהוי של כתב-יד לטיני וסיני ואשר בהם יכלו להגיע לאחוזי זיהוי גבוהים, בשנים האחרונות ניכרת התעניינות רבה בשפה הערבית.

בעוד שנדרשים ביצועי זמן אמת במערכות אשר מבצעות זיהוי כתב יד מקוון, גישות קונבנציונליות של זיהוי טקסט בדרך כלל ממתינות עד לסיום הכתיבה בכדי להתחיל בתהליך הזיהוי ובניתוח הטקסט. דחית תהליך הזיהוי עד לסיום הכתיבה אינו מאפשר ניצול טווח זמן הכתיבה, דבר אשר גורם להאטת המערכת ומונע מימוש של תכונות מתקדמות כגון השלמה ותיקון איות אוטומטי בזמן אמת.

במסגרת עבודה זו אנו מציגים שיטה המאפשרת סגמנטציה וזיהוי של כתב יד של טקסט הכתוב בשפה הערבית בזמן-אמת. אנו ממחישים את יכולת הביצוע של התהליכים אשר דורשים זמן עיבוד ממושך, תוך כדי ניצול זמן הכתיבה. הגישה אותה אנו מציגים מבצעת סגמנטציה תוך כדי התקדמות של משיכות עט. תהליך הסגמנטציה הינו מבוסס זיהוי, כלומר הוא משתמש בציונים אשר ניתנים על ידי המערכת לזיהוי אותיות לקטעים בתוך משיכת העט. בעבודה זו בנוסף אנו מציגים שיטה מהירה לזיהוי של אותיות הנדרשת על-ידי תהליך הסגמנטציה. מערכת זיהוי האותיות משתמשת בשיבוץ יעיל של התבניות למרחב האדוות (wavelets) אשר בו חיפוש התבניות הדומות ביותר הוא יעיל.

אנו מראים כי המידע אותו המערכת מצליחה לדלות תוך כדי התקדמות הכתיבה מאפשר צמצום משמעותי של גודל המילון הפוטנציאלי אשר בו יש לחפש בתהליך מאוחר יותר של זיהוי מדויק יותר של מילים בשיטה ההוליסטית.

בעקבות מחקר זה פורסמו שני מאמרים מדעיים אשר הוצגו בשני כנסים בינלאומיים בנושא זיהוי תבניות וזיהוי כתב יד.

אוניברסיטת תל-אביב

הפקולטה להנדסה ע"ש איבי ואלדר פליישמן

בית הספר לתארים מתקדמים ע"ש זנדמן-סליינר

סגמנטציה וזיהוי כתב יד מקוון תוך כדי

התקדמות של השפה הערבית

חיבור זה הוגש כעבודת מחקר לקראת התואר "מוסמך אוניברסיטה" בהנדסת חשמל
על ידי

ג'ורג' קור

העבודה נעשתה בבית הספר להנדסת חשמל
במחלקה למערכות

בהנחיית פרופ' דנה רון וד"ר ראיד סעאבנה

אייר תשע"ד

אוניברסיטת תל-אביב

הפקולטה להנדסה ע"ש איבי ואלדר פליישמן

בית הספר לתארים מתקדמים ע"ש זנדמן-סליינר

סגמנטציה וזיהוי כתב יד מקוון תוך כדי

התקדמות של השפה הערבית

חיבור זה הוגש כעבודת מחקר לקראת התואר "מוסמך אוניברסיטה" בהנדסת חשמל
על ידי

ג'ורג' קור

אייר תשע"ד