

תכונות מנוע החיפוש Fast

עפר דרורי

offerd@gmail.com

מבוא

מטרת מסמך לזה לסקור את תכונותיו של מנוע האחזור Fast בהתבסס על קריטריונים שונים המבוססים על מסמך הקריטריונים להשוואת מנועי אחזור גרסה 5.

המסמך מרכז קריטריונים רבים לצורך בדיקה והשוואה בין מוצרים שונים והוא מהווה כלי עזר לבחירה של מוצר לארגון על פי הצרכים הארגוניים שנקבעו מראש. הטבלה כוללת התייחסות למספר נושאים:

נתוני זיהוי של המוצר, החברה המפתחת והנציגות בארץ אם קיימת
תכונות המוצר, בקריטריונים השונים
דרישות פיתוח, ההנחה היא שמנוע החיפוש ישתלב במערכת גדולה יותר וחלק מהיכולות שלו יבואו לידי ביטוי במערכת זו באמצעות כלי פיתוח מתאימים
טכנולוגיה של המוצר, יכולת טכנולוגית, תמיכה וקישוריות למוצרים אחרים
ביצועים של המוצר, נושא חשוב לעיתים אף עולה בחשיבותו על תכונות מסוימות במוצר
מידע על הספק וניסיונו, בהתייחס לתצורות השונות

מידע נוסף על תכונות המנועים, אחזור טקסט ועוד נמצא באתר קבוצת העניין אחזור טקסט – SIGTRS
<http://www.sigtrs.org>

תודה לתמיר קסל שסיפק את המידע לגבי מנוע Fast.

פרטים מזהים

שם המוצר	FAST ESP
שם קודם של המוצר	
מספר גרסה נוכחית	5.3
שם החברה המפתחת	FAST, a Microsoft subsidiary
כתובת החברה	מיקרוסופט ישראל, הפנינה 2 רעננה
שם הנציגות בארץ	Microsoft Israel
כתובת הנציגות	
שם איש הקשר בארץ	תמיר קסל
טלפון איש קשר	054 663 2271
פקס איש קשר	
דואר אלקטרוני	tamirk@microsoft.com
אתר אינטרנט של המוצר	http://www.microsoft.com/enterprisesearch/en/us/fast.aspx

תכונות

המידע המצורף אינו מדריך רשמי של המוצר. המידע המוסמך הבלעדי הוא זה המסופק יחד עם המוצר

קריטריון	הסבר	התייחסות הספק
אחזור על מסמכי טקסט	תכונה בסיסית, בהתייחס לטקסט חופשי	נתמך
אחזור על מידע מפורמט	היכולת לבצע אחזור על שדות מידע רגילים מפורמטים	נתמך
אחזור לשדות מפורמטים בתוך מסמך טקסט	כמו תאריך, מחבר וכו'	נתמך
אופרטורים בולאנים	AND, OR, NOT, <, >, =, <=, >=, שימוש ב- () לביטויים מורכבים	נתמך
אופרטורים מטריים (מטריקה)	"מילה" AND "מילה שנייה" במרחק X מילים, באותו משפט, באותה פסקה וכו'	נתמך
אחזור לשפה עברית	הכוונה לטיפול מיוחד בשפה ולא לאחזור על בסיס ייצוג האותיות העבריות במאגר	בדיקת איות, מילים נרדפות, ניסוח, פיסוק ומילון התומך במורפולוגיה
אחזור לשפה אנגלית		נתמך
אחזור ל- 2 השפות במעורב	באותו מסמך	נתמך
אחזור לשפות נוספות	נא לפרט	תמיכה בלשנית פשוטה עבור למעלה מ-80 שפות ותמיכה בלשנית מתקדמת עבור 20 שפות (כגון: מורפולוגיה, פיסוק, ניסוח) כולל עברית, ערבית, אנגלית, רוסית ואחרות.
מילון מורפולוגי כחלק מהמוצר	האם קיים כזה ואם כן מהן תכונותיו	בדיקת איות, בדיקת ניסוח ופיסוק, מילים

נרדפות ומילון התומך במורפולוגיה		
תוסף "מורפיקס" זמין לשימוש והפתרון מוטמע בקרב מספר לקוחות	האם השילוב אפשרי, אם כן ציין איזה מילון (כמו מורפיקס של מלינגו) ואיזה גרסה	שילוב מילון מורפולוגי חיצוני
מסגרת עבודה וכלים עבור תזאורוס זמינים לשימוש (עבור שימוש עסקי ספציפי או בלשני כללי). תזאורוס זמין עבור אנגלית ועבור מספר שפות גלובאליות נפוצות נוספות	האם קיימת תשתית לשימוש בטבלאות תזאורוס חיצוניות, האם המוצר כולל תזאורוס משלו, אם כן לאיזה תחום ובאיזה שפה	תמיכה בטבלאות תזאורוס במוצר
ממשק משתמש מתוחכם עבור משתמשים עסקיים שמאפשר להם לנהל את המילונים	במידה ואנו רוצים לייצר את המילון לבד, האם יש תמיכה לשימוש בטבלה ריקה שתוזן ע"י המשתמש	ניהול תזאורוס במוצר
לא מגיע כתשתית מוכנה מראש אבל ניתן לבצע התאמות לצורך שימוש	האם קיימת תשתית במוצר לתזאורוס היררכי	ניהול תזאורוס היררכי
כן, ממשק משתמש כחלק מהפתרון וניתן להתאמות. בנוסף, ישנם רכיבי רשת, Web parts, הזמינים לצורך הצגת התוצאות במערכת SharePoint ומסגרת עבודה לבניה עצמית של רכיבי רשת.	האם המוצר כולל ממשק משתמש להצגת תוצאות החיפוש, אם כן באיזו שיטה: רק כותרות, תחילת מסמך, משפטיים רלוונטיים לחיפוש וכו' – יש לפרט	הצגת תוצאת החיפוש ע"י המוצר
נתמך	האם המוצר כולל ממשק משתמש לביצוע שאילתת החיפוש	הצגת שאילתת החיפוש ע"י המוצר
נתמך	האם המוצר כולל תמיכה בהדגשת מילים המקיימות את תנאי החיפוש במסמכים שאותרו	הדגשת מילים המקיימים את תנאי החיפוש
נתמך	יצירת סטים של תשובות וביצוע אחזור נוסף עליהם	ביצוע שאילתות על שאילתות
נתמך	האם אפשרי ואם כן האם גם בפורמטים שונים של המאגרים	אחזור על מספר מאגרים במקביל
נתמך	טיפול במצב מורכב בו כותרות המסמכים נמצאות בבסיס נתונים והמסמכים עצמם במערכת לניהול קבצים	אחזור על מסמכים מפוצלים
נתמך	האם קיים בלי הרחבות אוטומטיות	אחזור מדויק בלבד
נתמך	האם קיים	אחזור פונטי
נתמך	ראשיות, סופיות ואמצעיות אפשרות להחלפת תווים בסימן "*" (Wildcard)	הרחבות אחזור לחלקי מילים
FAST לא בנוי על בסיסי נתונים חיצוניים	האם נעשה במוצר עצמו, האם מתבסס על בסיסי נתונים חיצוניים, תיאור שיטת עבודה	ניהול אינדקסים במוצר
הרשאת הגישה לממשק הניהול	האם יש תמיכה להרשאות שונות על קטעים במאגר	ניהול הרשאות גישה לקטעים במאגר

האדמיניסטרטיבי והעסקי ניתנת להגדרה עבור משתמש. הרשאות אבטחת המידע משפיעות על ההרשאות המקוריות של מקור המידע (Security Trimming)		
נדרש גיבוי חיצוני אבל השיטה לעשות זאת מתועדת היטב	האם כולל מנגנון גיבוי ושחזור עצמי או שמתבסס על גיבוי ושחזור חיצוני של כל סביבת העבודה	גיבוי ושחזור במוצר
נתמך	האם קיים, מגבלות אם יש נא לציין	תמיכה באחזור מסמכי טקסט ASCII
נתמך		תמיכה במסמכי UNICODE Text
נתמך כמו גם תמיכה בחיפוש בתוך XML מבלי לבצע מיפוי מראש של האינדקס		תמיכה במסמכי XML כולל מסמכי אופיס שנשמרו כ- XML
נתמך, כולל הגרסאות האחרונות	להתייחס לאיזה גרסאות WORD, הכוונה לאחזור ישיר ממסמכי WORD ללא הסבתם	תמיכה באחזור מסמכי טקסט מסוג WORD
נתמך		אחזור מתוך מסמכי RTF
נתמך	האם קיים	תמיכה באחזור מסמכי טקסט מסוג HTML
	האם קיים, ציין מגבלות לגבי לוגית וחזותית	תמיכה באחזור מסמכי טקסט HTML בעברית לוגית וחזותית
נתמך	האם קיים, הכוונה לאחזור ישיר ממסמכי Excel	אחזור מתוך מסמכי Excel
נתמך		אחזור מתוך מסמכי Powerpoint
נתמך	האם קיים, ישירות מול קבצי PDF אקרובט	אחזור מתוך מסמכי PDF
נתמך	האם קיים, ישירות מול מסמכי Post Script	אחזור מתוך מסמכי PS
כל המסמכים זמינים בפורמט טקסט טהור וב-HTML	האם המוצר כולל תרגום פורמטים שונים לטקסט נקי, אם כן פרט אילו	תרגום פורמטים
לא. ניתן ליישם מורפיקס של מלינגו	האם קיים	אחזור לשורשים בשפה העברית והאנגלית
"דחיפה" ו"משיכה" של מסמכים זמינה באמצעות ממשק הפיתוח לתכנים-Content API	מתוך מערכת O.L., הכוונה למנגנון המאפשר הוספה מיידית של מסמך וביצוע אחזור מיידית למסמך במסגרת המוצר ולא להפעלת אצווה כל מספר שניות או דקות	אפשרות להוספת מסמכים בצורה מקוונת
ניתן לחילוף מתוך טבלאות ב-Data base	אחזור וניהול המידע כאשר הוא בטבלאות	תמיכה בטיפול בטבלאות
ל-FAST יש גמישות רבה בהשפעה על הדירוג,	האם המוצר כולל מנגנון לדירוג רשימת התוצאות, פרט אלו	דרוג (Ranking) תשובות

כולל באמצעות דירוג מקורות המידע, שדות, ערכים בשדות, חוקים לשוניים, לפי רענון הנתונים ויכולות רבות נוספות	אלגוריתמים קיימים לדירוג	
FAST מגיע עם יכולת חילוץ ישויות Out of the box ובנוסף סביבת עבודה לביצוע התאמות המאפשרת לבנות מחלצי ישויות המבוססים על קטגוריות או בהתאמה אישית	האם המוצר כולל מנגנוני סיווג כמו קטגוריזציה, אשכולות דינמיים ועוד. נא לפרט גם בהתייחס לשפה	מנגנונים לסיווג
משתמשים יראו רק את התוצאות שיש להם הרשאה עבורן. FAST יכול להתממשק עם כל LDAP התומך במודול אבטחה ועם מנגנונים נוספים שיש להם מתאמים ל-FAST	יכולת המוצר לתמוך במנגנוני אבטחה (קבצים או בסיסי נתונים) כך שהמשתמש יקבל את רשימת תוצאות החיפוש בהתבסס על ההרשאות שלו	אבטחת מידע
נתמך	יכולת המוצר לחלץ מידע על ישויות מתוך טקסט	חילוץ ישויות

דרישות פיתוח

התייחסות הספק	הסבר	קריטריון
.NET & Java	יכולת המוצר בתמיכה בסביבות פיתוח שונות SDK ל- Java .Net וכו'	תמיכה בסביבות פיתוח
נתמך	קיום SDK למגוון צרכים כמו: חיפוש, קלסיפיקציה, תחזוקת אינדקסים, תחזוקה כללית של מנהל המערכת, יכולות חיפוש מורחבות וכו'	קיום SDK
נתמך	קיום רכיבים לביצוע עיבודים שונים בתהליכי האינדוקס לדוגמא: ניתוח לשוני, קטגוריזציה, קלסיפיקציה, יצירת אשכולות וכו'.	תפיסת פיתוח תהליכי אינדוקס ואחזור

טכנולוגיה

קריטריון	הסבר	התייחסות הספק
תמיכה בשרת/לקוח	האם קיימת	נתמך
תמיכה בלקוח תחת חלונות XP, 2000, 98 וכו' דפדפנים לסוגיהם	התייחס לגרסאות התומכות בסביבות העבודה השונות, לגבי דפדפנים התייחס לגרסאות השונות, אם ידועות בעיות בנושא - נא ציין	תמיכה מובנית של הממשק ב-Explorer
תמיכה בשרתי UNIX, NT	האם קיימת	תומך ב- Windows, Linux, Unix
קישור לבסיסי נתונים מסוג SQLServer, אורקל	האם קיימת תמיכה למידע המאוחסן בבסיסי הנתונים הנ"ל, אם קיימת תמיכה בבסיסי נתונים מקומיים אחרים נא ציין	תומך בכל בסיסי הנתונים התואמים ל- (Oracle) DBC, MSSQL (וכד')
קישור ל- Exchange	האם קיים בכל הקשור לאחזור מידע המוטמע בשרתי Exchange	נתמך
קישור ל- Notes	האם קיים בכל הקשור לאחזור מידע המוטמע בשרתי Domino	נתמך
ממשקים מסוג Java, OLE, RPC, ACTIVEX	ציין אילו קיימים ואם קיימים נוספים נא ציין	ממשקים זמינים עבור JAVA ו-net. כמו כן תוצאות יכולות להיות מוחזרות כ-XML
ממשק API	האם קיים ממשק API המאפשר ביצוע של כל התכונות מתוך תוכנה אחרת חיצונית	נתמך
קישור לשרת אינטרנט IIS	האם קיים	תמיכה ב-Crawl עבור אתרי Internet/Intranet
קישור לשרת אינטרנט של אורקל	האם קיים	תמיכה ב-Crawl עבור אתרי Internet/Intranet
קישור לבסיס נתונים ב- MF כאשר אינדקסים נמצאים ב- MF או בשרת	הכוונה למצב בו המידע המיועד לאחזור מאוחסן בבסיס נתונים מרכזי ב- M.F. כדוגמת ADABAS והאינדקסים לאחזור נמצאים בשרת או ב- M.F. (נא ציין במה המוצר תומך)	לא נתמך
תמיכה בריבוי אינדקסים	מה המספר המקסימאלי של אינדקסים שניתן להקים ולנהל?	FAST כולל אינדקס מרכזי מנוהל אחד עם הרבה collections שהם מבוססים על חלוקה לוגית של האינדקס
אינדוקס ישירות מבסיס הנתונים	האם מתבצע אינדוקס אוטומטי עם עדכון בסיס הנתונים?	נתמך- תלוי בטריגר הרלבנטי ב-DB
שמירת האינדקס בבסיס הנתונים	האם יש אפשרות לכלול את	לא

	האינדקסים בבסיס הנתונים עצמו?	
נתמך ל-FAST יש הטמעות של מאות מיליוני מסמכים/רשומות	האם קיימת	תמיכה בבסיס נתונים מעל 10 מיליון מסמכים
נתמך. Linear Scalability on all aspects of the system including query, indexing and document processing	האם ניתן לתמוך בנפחי מסמכים / שאילתות גדולים ע"י הפעלת מספר שרתים במקביל	יכולת Scale Out
נתמך לחלק מה sub-processes	האם תהליכי המערכת בנויים לריבוי ליבות עיבוד ולריבוי נימים (Multithreading)	ניצול מעבדים מרובי ליבות (core)

ביצועים

התייחסות הספק	הסבר	קריטריון
זה תלוי בחומרה, בפונקציונאליות ובנתונים עצמם. ישנן מספר לקוחות FAST שיש להם דרישות indexing latency מאד נמוכות. יש ליצור קשר עם מיקרוסופט לפרטים ספציפיים	אינדוקס של 100,000 רשומות חדשות, נא ציין ביחס לשרת סטנדרטי (ציין חוזק השרת, מספר CPU וכו')	זמן טעינה של 100,000 מסמכים
זה תלוי בחומרה, בפונקציונאליות ובנתונים עצמם. ישנן מספר לקוחות FAST שיש להם דרישות indexing latency מאד נמוכות. יש ליצור קשר עם מיקרוסופט לפרטים ספציפיים	כני"ל לגבי 10,000	זמן טעינה של 10,000 מסמכים
זה תלוי בחומרה, בפונקציונאליות ובנתונים עצמם. ישנן מספר לקוחות FAST שיש להם דרישות indexing latency מאד נמוכות. יש ליצור קשר עם מיקרוסופט לפרטים ספציפיים	עפ"י מבחני ביצועים של המוצר (נא ציין המבחן)	זמן תגובה לחיפוש ביחס לגודל המאגר

קשר עם מיקרוסופט לפרטים ספציפיים		
זה תלוי בחומרה, בפונקציונאליות ובנתונים עצמם.	בזמן נתון ועם גידול המאגר, נא לצרף נתונים	נפח אינדקס ביחס לגודל המאגר

מידע על הספק וניסיונו

התייחסות הספק	הסבר	קריטריון
מיקרוסופט מפתחת פתרונות חיפוש מזה מספר עשורים		ניסיון בשנים בפיתוח מנוע אחזור טקסט
4 (כולל התקנות OEM)		מספר התקנות בארץ
מעל 3000		מספר התקנות בחו"ל
מספר מאות		כמות התקנות למאגרים מעל 10 מיליון מסמכים
כמעט כל 3000 הלקוחות		כמות התקנות מעל מיליון מסמכים
NA		כמות התקנות מעל 100,000 מסמכים
מידע חסוי		כמות התקנות בסביבת אינטרנט/אינטרה-נט
מידע חסוי	נא לציין את סוג בסיס הנתונים : אורקל, SQL-Server וכו'	כמות התקנות המשלבת בסיס נתונים ארגוני על PC
מידע חסוי	נא לציין את סוג בסיס הנתונים : אדבס, DB2 וכו'	כמות התקנות המשלבת בסיס נתונים ארגוני על MF
מידע חסוי		כמות התקנות עם Exchange
מידע חסוי		כמות התקנות עם Notes
3 רמות תמיכה מטעם צוות התמיכה הגלובאלי, זמינות 24/7 או 9*5, באמצעות אימייל, כרטיס ותמיכה טלפונית	נא לפרט : תמיכה טלפונית או אחרת	סוגי תמיכה
שנתי		פרק זמן בין גרסה לגרסה
מיקרוסופט הכריזה לאחרונה על Roadmap רב-שנתי FAST עבור מוצרי	האם מוגבלת בזמן, האם מרכז עיסוק החברה בתחום, תוכניות לעתיד בהקשר למוצר	מחויבות החברה למוצר
מידע חסוי- ראה באתר אינטרנט	עפ"י הקריטריונים השונים : לקוחות עם בסיס נתונים אדבס, DB2 אחר לקוחות עם בסיס נתונים אורקל, SQL-SERVER וכו'.	רשימת לקוחות
עד 8 שנים		מספר שנים שהמוצר הנוכחי

		מותקן אצל לקוחות
מידע חסוי	אם המוצר השתתף במכרזים או בחירה אחרת ונבחר, נא ציין את הגורם הבוחר, שנת הבחירה ואת המוצרים שמולם עמד לתחרות	השתתפות במכרזים או בחירות קודמות

מחיר

התייחסות הספק	הסבר	קריטריון
צור קשר עם מיקרוסופט	נא להתייחס לגרסאות שונות של המוצר (אם קיימות) בהתייחס למספר שרתים, לאתר, שיטות רישוי שונות	מחיר המוצר
צור קשר עם מיקרוסופט לקבלת הצעת מחיר	נא לציין את תקופת האחריות	מחיר אחזקה לשנה לאחר תקופת האחריות
צור קשר עם מיקרוסופט	כולל משך זמן האחריות, ניתן לציין מחיר אפס	אחריות