נספח 5ב׳ |

מסמך הבהרות משלים למבחן היכולת

בהמשך להוראות נספח 5ב׳ ולמענה לשאלות ההבהרה, מובהרים להלן תנאי ביצוע מבחן היכולת. ההבהרות שלהלן משלימות את הוראות המכרז בנושאים המפורטים בהן; יתר הדרישות והוראות הניקוד נותרות לפי מסמכי המכרז.

א. תכולת מבחן היכולת ושדות החילוץ

מבחן היכולת יכלול ארבעה סוגי מסמכים: היתר בנייה מילולי, גרמושקה, תצ״ר וטבלת זכויות והוראות בנייה. בכל סוג ייבחנו ארבעת השדות המוגדרים בסעיף 6 לנספח 5ב׳ (16 שדות בסך הכול). אין בכך כדי לשנות את דרישות החילוץ הרחבות החלות על המערכת נשוא המכרז.

סוג מסמך	השדות הנכללים במבחן
היתר בנייה מילולי (permit)	setl, gush, parcel, bld_num
גרמושקה (garmushka)	gush, parcel, bld_num, primary_use
תצ״ר (tazar)	setl, gush, parcel, tazar_id
טבלת זכויות והוראות בנייה (mavat)	plan_id, setl, primary_use, lot_size

ב. סביבת הביצוע ואופן קבלת המסמכים

מבחן היכולת יבוצע בסביבת העבודה ובענן של המציע, על חשבונו, ולא נדרש לבצעו בענן של המזמין או בסביבת נימבוס.

מסמכי המבחן יועמדו לרשות המציעים במועד המבחן באמצעות קישור ל־Google Drive. על המציע להעמיד מראש את התשתיות, החיבור לרשת ויכולת הורדת המסמכים. כל המציעים ייבחנו על אותם מסמכים.

ג. שימוש במודלים ובשירותי עיבוד חיצוניים

במבחן היכולת ניתן להשתמש במודלים מסחריים ובשירותי AI/OCR בהתאם לפתרון שבחר המציע, בכפוף להוראות המכרז, לסיווג המידע ולדרישות שמירת מידע, סודיות וזכויות שימוש.

ההיתר להפעלת שירותים בסביבת המציע לצורך המבחן אינו מהווה אישור לשימוש באותם שירותים בסביבת הייצור בנימבוס, ואינו גורע מדרישות האישור החלות על מוצרים המוצעים במסגרת ההצעה.

ד. היערכות לפתרון והיעדר התערבות אנושית

מבחן היכולת יבוצע באמצעות פתרון שהוכן מראש. לא יידרשו ולא יתאפשרו פיתוח או התאמת יכולות חדשות במהלך המבחן.

הפקת התוצרים המנוקדים תיעשה באופן אוטומטי; לא תתאפשר התערבות אנושית, לרבות תיקון ידני של ערכים שחולצו או נורמלו.

ה. מספר המסמכים, איכותם ואופן ההרצה

לכל אחד מארבעת סוגי המסמכים תועמד אצווה נפרדת של עד עשרה מסמכים. האצוות יורצו בזו אחר זו ולא במקביל. לכל אצווה תבוצע ריצה אחת בלבד לצורך מדידה וניקוד, ולא תתאפשר בחירת התוצר המיטבי מבין ריצות חוזרות.

הקלט עשוי לכלול מסמכים בפורמטים שונים ומסמכים מרובי עמודים, ובכלל זה טבלת הוראות וזכויות בניה (טבלה 5) מרובה עמודים. חלק מהסריקות עשויות להיות מסובבות או ללא מטא־דאטה המציין אוריינטציה.

ו. קובצי תוצר וחלוקתם לפי מסמך

עבור כל מסמך מקור יוגשו קובץ CSV אחד וקובץ JSON אחד. עבור כל מסמך שניתן לשייכו מרחבית יוגש נוסף על כך קובץ GeoJSON נפרד. אין חובה להגיש קובץ GeoJSON מאוחד לכל סוג מסמך.

שמות קובצי התוצר ייגזרו משם קובץ המקור, ללא סיומת המקור ובצירוף סיומת התוצר המתאימה.

קובצי CSV יישמרו בקידוד UTF-8 בעברית, בהתאם לנספח 5ב

ז. נרמול שמות יישובים וערכים מרחביים

ערך השדה setl בעמודת normalized_value ייקבע בהתאם לשם היישוב כפי שהוא מופיע בשדה HEB_NAME בשכבת המידע Setl_lamas שסופקה למציעים, אף אם במסמך המקור מופיע שם היישוב בכתיב שונה.

נרמול ערכי גוש וחלקה יבוצע בהתאם לכללי הנרמול הקבועים בסעיף 8 לנספח 5ב.'

הבהרה לעניין חילוץ ונרמול השדה :primary_use

במסמכי גרמושקה (garmushka) המידע הנדרש לחילוץ עבור השדה primary_use עשוי להופיע תחת כותרות שונות, כגון "שימוש עיקרי", "מטרת הבניין" או כותרות אחרות בעלות משמעות דומה, בהתאם למבנה המסמך.

על המציע לחלץ את הערך הרלוונטי מתוך אזורי העניין כפי שמסומן במסגרות אדומות בדוגמאות המסמכים שפורסמו בקישור המצורף למסמך ההבהרות, גם כאשר כותרת השדה במסמך המקור שונה משם השדה המוגדר בפלט.

מובהר כי במסמכי גרמושקה (garmushka) ובמסמכי טבלת זכויות והוראות בנייה (mavat), עבור השדה primary_use הערך המנורמל (normalized_value) יהיה זהה לערך שחולץ (extracted_value) ללא שינוי, המרה, האחדת כתיב או מיפוי לערך אחר.

בכל המקרים שם השדה בקובצי הפלט יהיה primary_use בהתאם להוראות נספח 5ב.'

יתר כללי הנרמול יישארו בהתאם להוראות סעיף 8 לנספח 5ב.'

```
{
  "field_name": "primary_use",
  "extracted_value": "מגורים ב׳",
  "normalized_value": "מגורים ב׳"
}
```

ח. טיפול בערך שלא ניתן לחלץ

במקרה שערך אינו מופיע במסמך, או שאינו ניתן לזיהוי בוודאות מספקת, יש לפעול בהתאם לנספח 5ב׳. יש לייצג ערך חסר כ־"null" בקובץ CSV וכ־null בקובץ JSON. אין להשלים ערכים על דרך השערה.

ט. חילוץ מטבלת זכויות והוראות בנייה

במבחן היכולת, כאשר טבלה 5 כוללת מספר שורות, יש לחלץ את ארבעת השדות המוגדרים במבחן מן השורה הראשונה העליונה בלבד, לפי הדוגמה המסומנת בנספח. מובהר כי הוראה זו חלה על מבחן היכולת בלבד ואינה מצמצמת את דרישת חילוץ כלל השדות והערכים בטבלה 5 בשלב פיתוח המערכת.

י. הפקת GeoJSON ושיוך מרחבי

לצורך מבחן היכולת יחולו הוראות סעיף 5.1 לנספח ביחס למבנה FeatureCollection, לשיוך מרחבי ולמערכת הקואורדינטות EPSG:4326.

כאשר במסמך מופיעים מספר גושים, לצורך שיוך מרחבי יש להתייחס לגוש הראשון; ביחס לריבוי חלקות חלים הכללים שבנספח

לצורך הצגת מסמכי טבלת זכויות והוראות בנייה (mavat) מותר להשתמש בנקודת היישוב (Point) המופיעה בשכבת Setl_lamas שסופקה למציעים, והיא תיחשב לצורך המבחן כנקודת ה־Centroid הנדרשת.

יא. תחילת מדידת זמן העיבוד וסיומה

המסמכים יועלו למערכת לפני תחילת המדידה. מדידת זמן העיבוד תחל בלחיצה על פעולת "הרצה" ותסתיים לאחר שכל מסמכי האצווה הגיעו לתוצאה סופית והופקו עבורם קובצי ה־CSV וה־JSON הנדרשים, בהתאם לכללי נספח 5ב.

זמן העלאת מכונות והמסמכים לפני ההרצה, זמן הפקת קובצי GeoJSON וזמן הצגת התוצרים על גבי המפה אינם נכללים ברכיב זמן העיבוד. עיבוד המסמכים שחל לאחר הפעלת האצווה נכלל במדידה.

יב. סף הדיוק לציון רכיב הזמן

חישוב זמן הייחוס לכל סוג מסמך ייעשה על בסיס הזמן הקצר ביותר שהושג בקרב מציעים אשר עמדו בדרישת דיוק מינימלי של 80% באותו סוג מסמך, כקבוע בסעיפים הנספח. יתר הוראות הציון והמשקלים הן כמפורט במסמכי המכרז.

יג. דוגמאות מסמכים, אזורי עניין, נתוני אמת ושכבות הנרמול (ישובים, גוש וחלקה)

לצורך היערכות למבחן היכולת, מצורף קישור לדוגמאות מסמכים מארבעת סוגי המסמכים הנכללים במבחן, פלט גייסון נדרש עם נתוני אמת ושכבות הנרמול לטובת המבחן.

קישור לדוגמאות המסמכים:

https://drive.google.com/drive/folders/1F2iG1n0o031ngKwORHInPg7q2eEQK7ln?usp=sharing

בדוגמאות המסמכים, סומנו במסגרות אדומות אזורי העניין שבהם מופיעים הערכים והשדות הנדרשים לחילוץ, בהתאם להוראות נספח 5ב.'

מובהר כי המסגרות האדומות נועדו להמחשה ולהכוונת המציעים בלבד. מסמכי מבחן היכולת עצמם יימסרו ללא סימון המסגרות האדומות, ועשויים לכלול מספר עמודים לכל מסמך .על המציע לזהות באופן עצמאי את אזורי העניין הרלוונטיים, לאתר ולחלץ מהם את הערכים הנדרשים, ללא הסתמכות על סימון מקדים של אזורי החילוץ.

הדוגמאות נועדו להמחיש את מיקום השדות ואת אופן הופעתם במסמכים, ואינן מגבילות את מיקום השדות או את אופן הצגתם במסמכי המבחן.

יתר הוראות נספח 5ב', לרבות דרישות החילוץ, הנרמול והגשת התוצרים, יישארו ללא שינוי.

יד. הצגת תוצאות על גבי מפה

הצגת תוצאות החילוץ על גבי רכיב מפתי תבוצע באמצעות GOVMAP API, בהתאם לנספח 5ב׳. לצורך קבלת טוקן על המציע להירשם למערכת GOVMAP ולהגיש בקשה חדשה דרך אזור ״ניהול API״.

בבקשה יש לציין במפורש: ״עבור מכרז לחילוץ מידע ממסמכים סרוקים״. טוקן שאושר יהיה זמין במערכת GOVMAP עצמה.

מומלץ למציעים להגיש את הבקשה לקבלת הטוקן כבר כעת ובהקדם האפשרי, ולא להמתין למועד מבחן היכולת, כדי לאפשר היערכות מוקדמת לבדיקת הממשק ולהצגת התוצרים המרחביים. על המציע להיערך מבעוד מועד לשימוש ב־GOVMAP API במסגרת מבחן היכולת.

טו. מבנה פלט JSON ודוגמאות

כמפורט בסעיפי נספח 5ב', יוגש לכל מסמך קובץ JSON תקין ונפרד.

אובייקט השורש יכלול את המאפיינים הבאים:

- file_name – שם הקובץ ללא סיומת.
- document_type – אחד מהערכים permit, garmushka, tazar, mavat.
- fields – מערך של ארבעת השדות הנדרשים לחילוץ

לכל שדה יופיעו שלושת המאפיינים הבאים:

- field_name – שם השדה.
- extracted_value – הערך כפי שחולץ ממסמך המקור.
- normalized_value – הערך לאחר נרמול בהתאם לכללי הנספח

- ערך שלא חולץ יוצג כ־ null תקני ב־ JSON ללא מרכאות.
- ערכי הדוגמאות שלהלן ממחישים את מבנה הפלט הנדרש למבחן היכולת

דוגמה 1 – היתר בנייה מילולי | P1.json

```
{{
  "file_name": "P1",
  "document_type": "permit",
  "fields": [
    {
      "field_name": "setl",
      "extracted_value": "תל אביב",
      "normalized_value": "תל אביב -יפו"
    },
    {
      "field_name": "gush",
      "extracted_value": "123",
      "normalized_value": "123"
    },
    {
      "field_name": "parcel",
      "extracted_value": "4567",
      "normalized_value": "4567"
    },
    {
      "field_name": "bld_num",
      "extracted_value": "123/45",
      "normalized_value": "123_45"
    }
  ]
}
```

דוגמה 2 – גרמושקה | G2.json

```
{
  "file_name": "G2",
  "document_type": "garmushka",
  "fields": [
    {
      "field_name": "gush",
      "extracted_value": "123",
      "normalized_value": "123"
    },
    {
      "field_name": "parcel",
```

```
    "extracted_value": "12,56,34",
    "normalized_value": "12,34,56"
   },
   {
    "field_name": "bld_num",
    "extracted_value": null,
    "normalized_value": null
   },
   {
    "field_name": "primary_use",
    "extracted_value": "מסחרי",
    "normalized_value": "מסחרי"
   }
  ]
}
```

דוגמה 3 – תצ״ר | T3.json

```
{
 "file_name": "T3",
 "document_type": "tazar",
 "fields": [
   {
    "field_name": "setl",
    "extracted_value": "תל אביב",
    "normalized_value": "תל אביב -יפו"
   },
   {
    "field_name": "gush",
    "extracted_value": "123",
    "normalized_value": "123"
   },
   {
    "field_name": "parcel",
    "extracted_value": "4567",
    "normalized_value": "4567"
   },
   {
    "field_name": "tazar_id",
    "extracted_value": "123/045",
    "normalized_value": "123_045"
   }
  ]
}
```

דוגמה 4 – טבלת זכויות והוראות בנייה | M4.json

```
{
 "file_name": "M4",
 "document_type": "mavat",
 "fields": [
   {
    "field_name": "plan_id",
    "extracted_value": "123-045",
    "normalized_value": "123_045"
   },
   {
    "field_name": "setl",
    "extracted_value": "תל אביב",
    "normalized_value": "תל אביב -יפו"
   },
   {
    "field_name": "primary_use",
    "extracted_value": "מגורים ב׳",
    "normalized_value": "מגורים ב׳"
   },
   {
    "field_name": "lot_size",
    "extracted_value": "123",
    "normalized_value": "123"
   }
```

```
  ]
}
```

טז. דוגמה להפקת GeoJSON

להמחשת סעיפי הנספח, להלן דוגמת GeoJSON במבנה FeatureCollection הכולל Feature מסוג Point. הקואורדינטות שבדוגמה הן להמחשה בלבד ואינן מחליפות שיוך לגוש/חלקה או חישוב מיקום בהתאם לנספח. התוצר יוצג על גבי רכיב מפתי המבוסס על GOVMAP API.

```
{
  "type": "FeatureCollection",
  "features": [
    {
      "type": "Feature",
      "geometry": {
        "type": "Point",
        "coordinates": [
          34.7818,
          32.0853
        ]
      },
      "properties": {
        "file_name": "P1",
        "document_type": "permit",
        "setl": "תל אביב -יפו",
        "gush": "123",
        "parcel": "4567",
        "bld_num": "123_45"
      }
    }
  ]
}
```

בכל נושא שלא הוסדר במפורש במסמך זה ימשיכו, לחול הוראות מסמכי המכרז ומעני ההבהרה הרשמיים, כפי שפורסמו.