

UTF8-ו JOOMLA

מה זה יוניקוד (Unicode) ו-UTF8

יוניקוד (Unicode) הוא תקן בינלאומי לייצוג טקסט במערכות מחשב. התקן מגדיר מערכת תווים המקיפה את כל מערכות הכתב הנמצאות כיום בשימוש פעיל בשפות העולם. לפני שהחל השימוש בתקן זה, התקן המקובל לייצוג טקסט היה ASCII. תקן זה הכיל 256 תווים: 128 תווי ASCII ו-128 תווים אחרים לפי האזור, כך שאותו מספר בעצם היה יכול להיות אותיות שונות. למשל, המספר 244 הכיל גם את האות à בקידוד של מערב אירופה וגם את האות א בקידוד העברי. מי שעייין בטקסט ראה את האות בהתאם לקידוד בו הוא צפה. אם היה בוחר לצפות בקידוד מערב אירופה היה רואה את האות à, אם היה בוחר לצפות בקידוד עברי, היה רואה את אותו טקסט עם האות א. דבר זה גרם לבעיות בהעברה של טקסט ממערכות שונות. בשנת 1990 התכנסו חברות תוכנה ותקשורת עולמיות כדי לפתור את הבעיה וכך החלה התקינה של יוניקוד. כיום, השימוש ביוניקוד מאפשר לייצג כל אות, בכל שפה, במספר ייחודי לה, ולכן ניתן לכתוב באותו הטקסט אותיות שפעם "יישוב" על אותו מספר.

UTF-8 הוא למעשה שם של צורת דחיסה עבור יוניקוד המשתמשת ב-8 תווים: **Unicode Transformation Format** 8 bit. כיום יש גם דחיסה ב-16 תווים וב-32 תווים.

גרסת ג'ומלה 1.5 כבר מוכנה לעבודה בקידוד UTF-8, אולם הגרסאות הקודמות לה אינן פועלות כברירת מחדל ביוניקוד. מדריך זה מסכם את פעולות ההמרה והעבודה של מערכת ג'ומלה בקידוד UTF-8, עבור גרסאות 1.0.X.

שינוי קובץ database.php

ראשית, יש לבצע שינוי באחד מקבצי המערכת, קובץ **database.php** הנמצא בתיקיית **includes**. יש לפתוח קובץ זה ולמצוא את חלק הקוד הבא, בשורה 102:

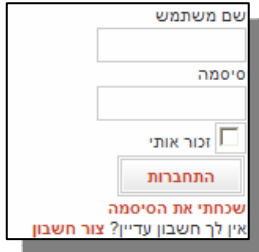
```
}  
$this->_table_prefix = $table_prefix;  
//@mysql_query("SET NAMES 'utf8'", $this->_resource); ←  
$this->_ticker = 0;  
$this->_log = array();  
}
```

בחלק קוד זה, שורה 102 מסומנת כהערה (על ידי //) יש להוריד את סימן ההערה, את שני הסלשים, כך שהקוד יראה כך:

```
}  
$this->_table_prefix = $table_prefix;  
@mysql_query("SET NAMES 'utf8'", $this->_resource); ←  
$this->_ticker = 0;  
$this->_log = array();  
}
```

שינוי קובץ השפה

אם באתר, עבור עברית, הותקן קובץ שפה אשר קודד מראש ל-UTF8 אין צורך לבצע כל שינוי. אולם, אם קובץ השפה שהותקן, הוא בקידוד ASCII (windows-1255) האתר יראה בג'יבריש, יש לשנות את הקידוד ל-UTF8.



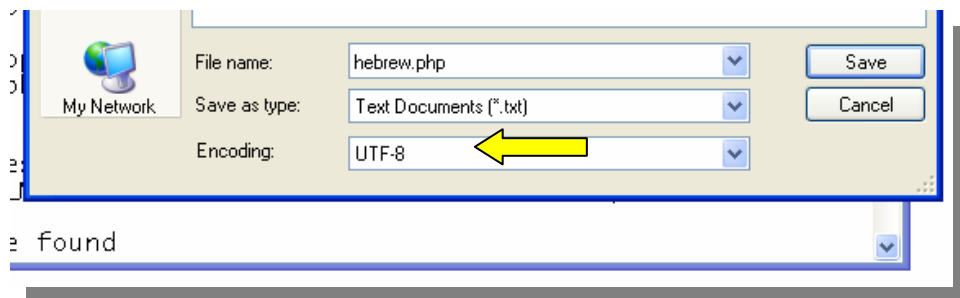
ניתן לזהות בקלות אם קובץ השפה תקין: אם הודעות המערכת יופיעו בצורה לא תקינה סימן שיש בעיה. למשל, שם משתמש, סיסמה, זכור אותי התחברות וכו' – יופיעו בג'יבריש או סימני שאלה.

קבצי השפה נמצאים בתיקיית **language**.

ראשית יש לפתוח את הקובץ **hebrew.xml** ולוודא כי הוא מורה על קידוד UTF-8 בשורה הראשונה:

```
<?xml version="1.0" encoding="utf-8"?>
<mosinstall type="language">
  <name>Hebrew</name>
```

לאחר מכן, יש לפתוח את שני הקבצים הנוספים בתיקיית השפה: **hebrew.php**, **hebrew.ignore.php** באמצעות NotePad. לכל אחד מהקבצים יש לבצע "שמירה בשם" ולשמור אותו תחת אותו שם, באותה תיקייה אולם לבחור בתחתית מסך השמירה בקידוד Unicode:

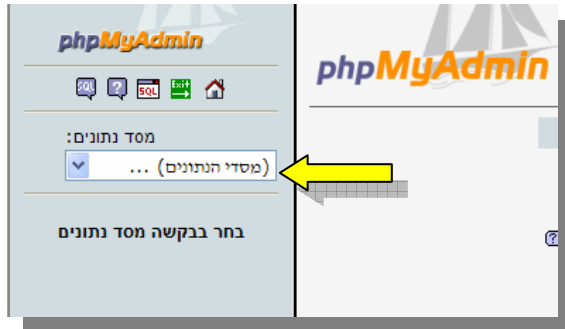


הערה: רצוי לא להשתמש בתוכנות עריכת HTML או PHP כגון dreamweaver, ולבצע את הפעולה באמצעות NotePad. תוכנות כאלו עלולות להוסיף ביט נוסף בתחילת הקובץ (נקרא BOM), דבר שיגרום לשגיאה בהמשך.

פעולה זו של הפיכת קידוד קובץ השפה, תהיה נכונה עבור קבצי שפה נוספים, למשל קבצי שפה של רכיבים שהותקנו באתר ויש להם קובץ שפה שתורגם לעברית.

הכנת בסיס הנתונים עבור UTF

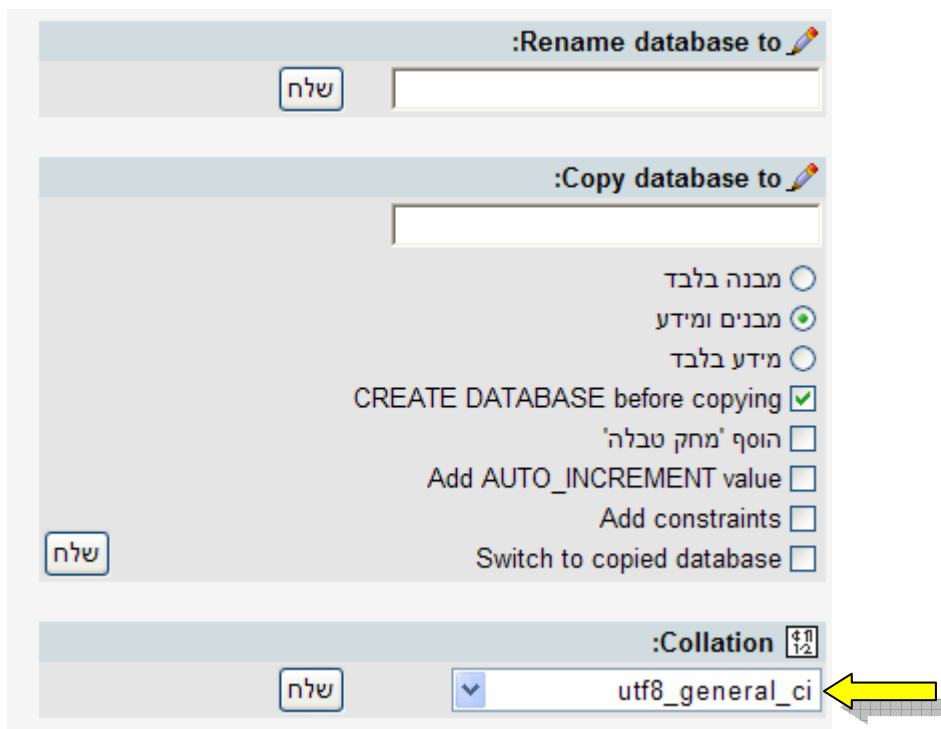
לפני התקנת האתר, בעת יצירת בסיס הנתונים יש לוודא כי הוא בקידוד UTF-8. בפאנל הניהול של חשבון האירוח (למשל בפאנל מסוג direct admin) ניתן ליצור בסיס נתונים בצורה מאד קלה, אולם בשלב היצירה בפאנל זה, לא ניתן לציין את קידוד בסיס הנתונים והוא ייווצר תחת הגדרות ברירת המחדל של השרת. ניתן להיכנס ל-phpMyAdmin ולשנות את קידוד בסיס הנתונים שכבר נוצר. יש לבחור את בסיס הנתונים הרצוי:



לאחר שנבחר בסיס הנתונים, יש להיכנס ל"פעולות" שבסרגל הניווט העליון

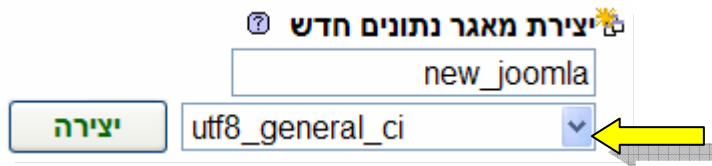


בתחתית החלון שיפתח, תחת Collation יש לבחור את הקידוד: utf8_general_ci



כעת בסיס הנתונים מוכן להתקנה.

בעת יצירת בסיס נתונים חדש ב-phpMyAdmin ולא דרך ממשק ניהול אחר, ניתן לציין את הקידוד כבר בעת יצירת בסיס הנתונים.



התאמת התבנית לקידוד UTF-8

יש להתאים את תבנית העיצוב המותקנת באתר לקידוד UTF-8. בתיקיית התבנית יש לפתוח את קובץ `templateDetails.xml` ולוודא כי בשורה הראשונה מצוין שקידוד התבנית הוא utf-8:

```
<?xml version="1.0" encoding="utf-8"?>
```

יש לפתוח גם את קובץ `index.php` של התבנית ולשמור אותו תחת קידוד utf-8. ניתן לבצע זאת באמצעות NotePad (כפי שהוסבר עבור קובץ השפה). רצוי לוודא שבקובץ זה לא קיימת שורה בתוך הקוד המורה על קידוד אחר:

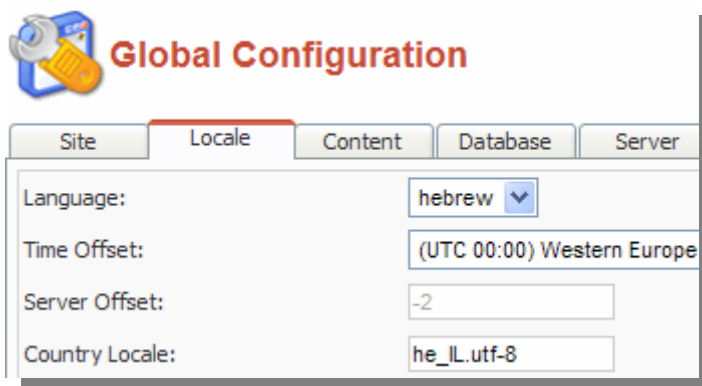
```
<meta http-equiv="Content-Type" content="text/html; charset=windows-1255" />
```

שורה כזו עלולה להופיע בין תגי `<head>`. אם היא מופיעה יש למחוק אותה.

כעת, לאחר ששונה קובץ `database.php`, הותאמו בסיס הנתונים, קובץ השפה והתבנית – האתר יכול לפעול תחת קידוד utf-8.

שינוי הגדרות אזוריות לאחר התקנה

לאחר ההתקנה, כדי להציג תאריכים באופן תקין ("יום ראשון" במקום "Sunday" ...), יש לשנות את ההגדרות האזוריות תחת לשונית Local ב-Global Configuration. יש לקבוע בשדה `county local` את הערך: `he_IL.utf-8`:



המרת אתר שהותקן על קידוד ASCII ל-UTF8

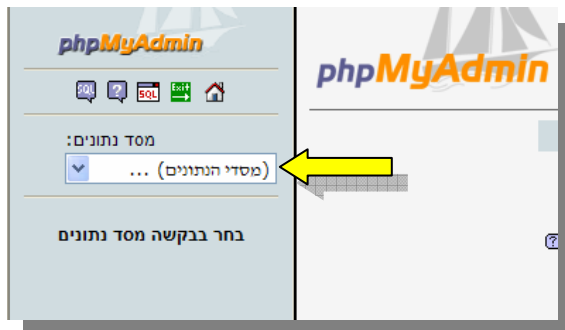
שים לב:

- יש לבצע גיבוי מלא של האתר ושל בסיס הנתונים לפני ביצוע פעולה זו.
- הפעולה היא על אחריות המשתמש.

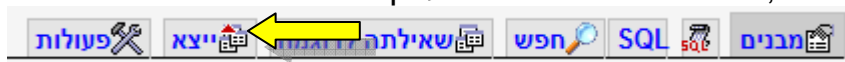
1. יש לבצע התאמה של קובץ השפה כפי שהוסבר קודם לכן.
2. יש לבצע התאמה של קבצי שפה עבור רכיבים, כפי שבוצע עבור קובץ השפה.
3. יש לבצע התאמה של התבנית כפי שהוסבר קודם לכן.

התאמת בסיס הנתונים לאתר קיים

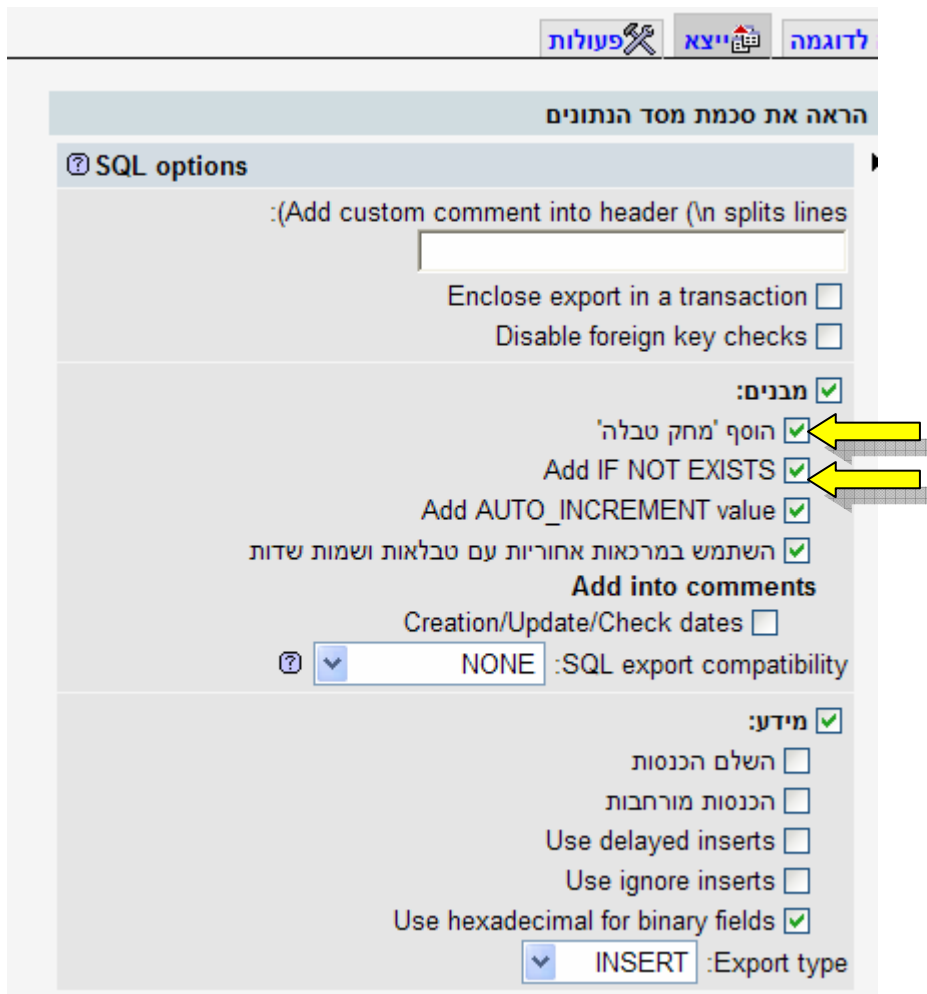
ב-PhpMyAdmin ניתן לייצא את בסיס הנתונים לקובץ SQL.
יש לבחור בסיס הנתונים הרצוי



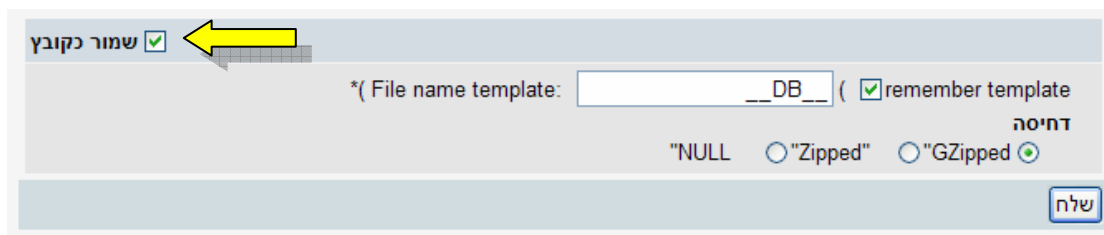
לאחר שנבחר בסיס הנתונים, יש לבחור בסרגל הניווט העליון ב"ייצא".



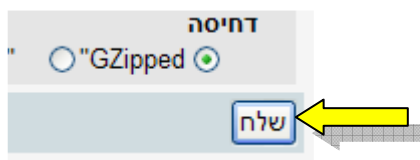
במסך הייצוא יש לסמן את האפשרויות "הוסף 'מחק טבלה'" ואת " Add "AUTO_INCREMENT value".



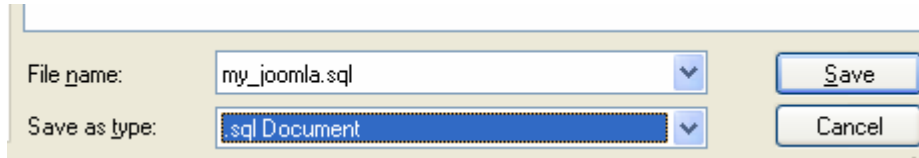
בתחתית מסך הייצוא יש לסמן את "שמור כקובץ"



ולבצע שמירה



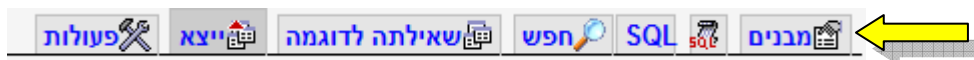
קובץ מסוג SQL ישמר אל המחשב



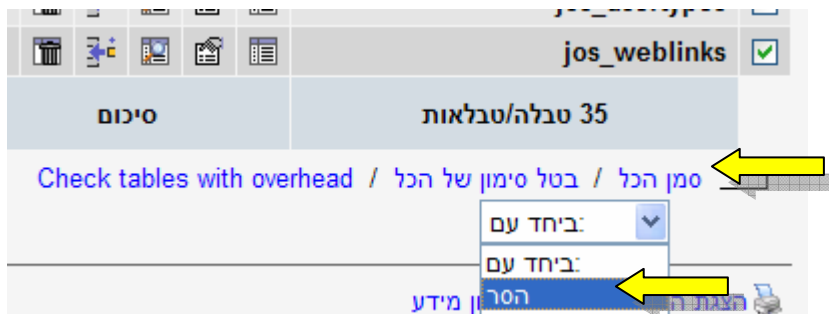
כעת יש לגשת אל קובץ זה, לפתוח אותו באמצעות NotePad ולבצע שמירה חוזרת ולשמור בקידוד Unicode (כפי שהוסבר עבור קבצי השפה, "שמירה בשם"...)...

כעת בסיס הנתונים שבקובץ ה-SQL הוסב ל-UTF8 ויש להתאים את בסיס הנתונים ל-UTF8.

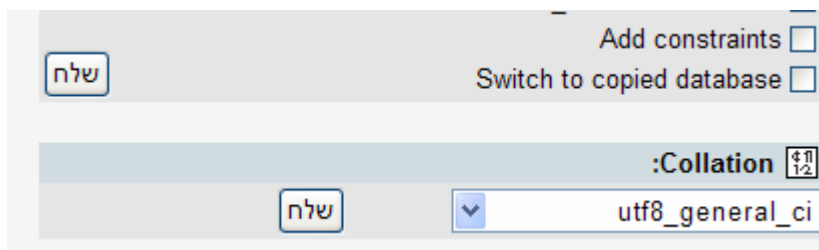
ראשית, ניתן למחוק את כל הטבלאות מבסיס הנתונים. בסרגל הניווט העליון יש לעבור ל"מבנים"



במסך שיופיעו, יוצגו כל הטבלאות של בסיס הנתונים. בתחתית רשימת הטבלאות יופיע הכיתוב "סמן הכל". לחיצה עליו תסמן את כל הטבלאות. כעת, למחיקת הטבלאות, יש לבחור מתוך תיבת הבחירה שמתחת, את האפשרות "הסר". **שים לב: פעולה זו תמחק את כל טבלאות ואת נתוני האתר! וודא שביצעת גיבוי לבסיס הנתונים ובצע פעולה זו על אחריותך בלבד!**



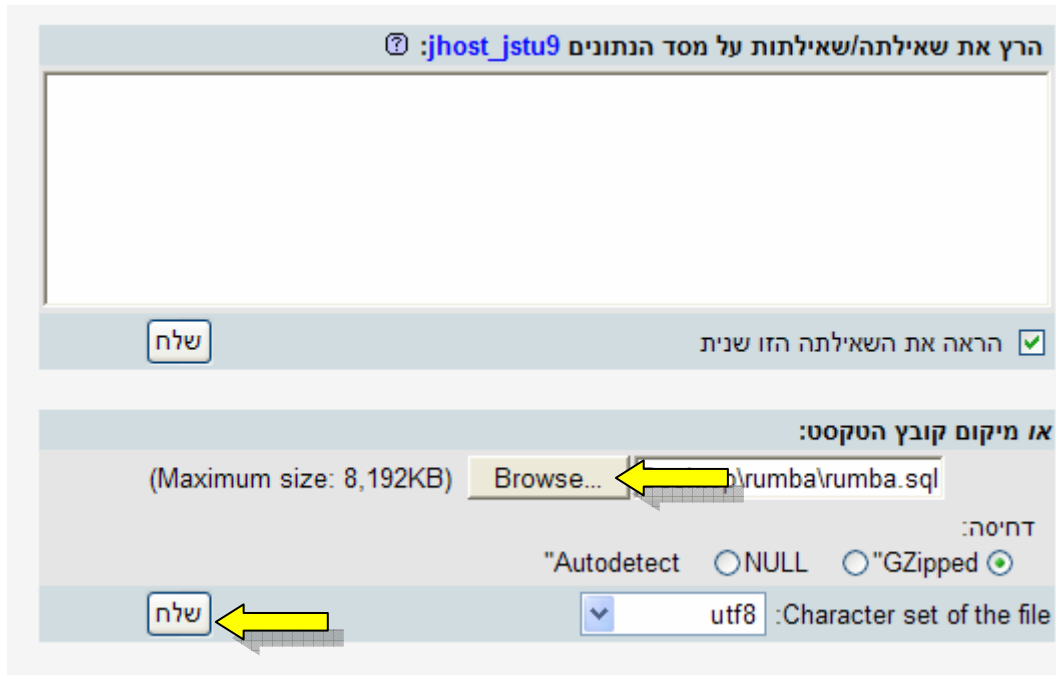
לאחר שנמחקו כל הטבלאות, ניתן להמיר את קידוד בסיס הנתונים ל-UTF8, כפי שהוסבר קודם לכן, עבור בסיס הנתונים חדש. יש לעבור ל"פעולות", ולבחור במסך זה, בקידוד: utf8_general_ci



כעת, בסיס הנתונים הותאם לקידוד הרצוי, וניתן להעלות אליו את קובץ ה-SQL שנשמר על המחשב והותאם לקידוד. יש לעבור בסרגל הניווט העליון למסך "SQL".



במסך זה, יש לבחור את קובץ ה-SQL שנשמר על המחשב ועבר הסבה וללחוץ על "שלח" כדי להעלות אותו אל בסיס הנתונים.



לאחר שהסתיים ייבוא הקובץ, וכל הפעולות הנוספות בוצעו – האתר הותאם ל-UTF8!