

مخيم غوغل للتدريب

تطوير البرمجيات بالعربية

المسائل والطرائق

طه زروقي

المدرسة الوطنية العليا للمعلوماتية الجزائر

Taha.zerrouki@gmail

الإشكالية

- أهمية تطوير البرمجيات بالعربية
- المسائل التقنية المتعلقة

أهمية تطوير البرمجيات بالعربية

أهم المسائل

■ التعامل مع الترميز

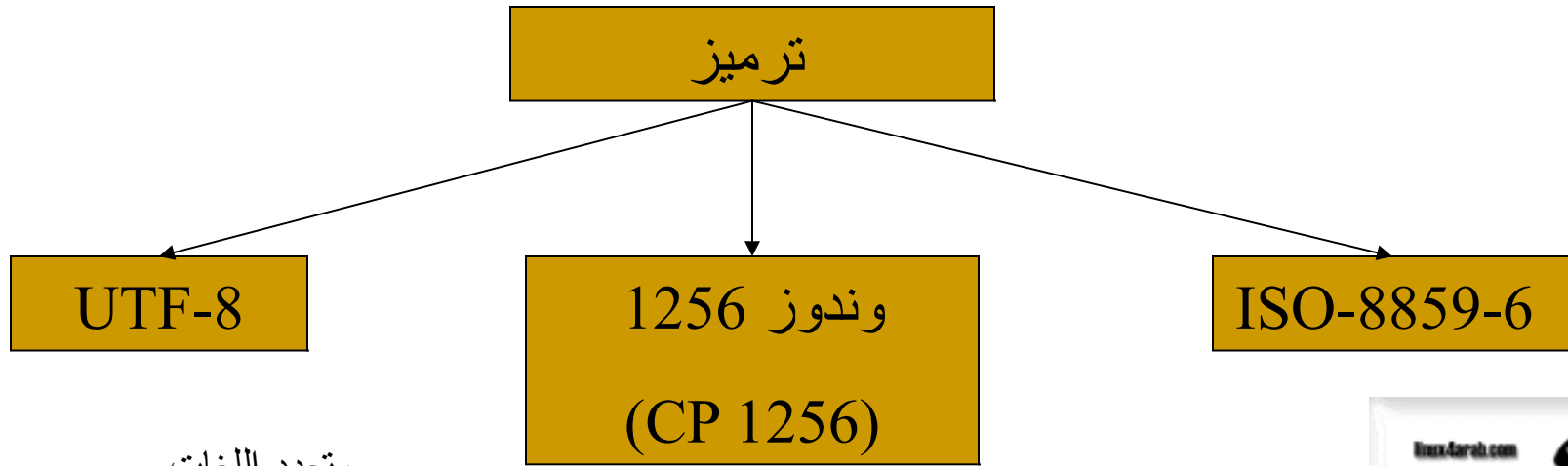
- اختيار الترميز
- ترميز صفحات الويب
- إعداد النظام
- التعامل البرمجي
- اليونيكود

■ عرض النص

- الكتابة من اليمين إلى اليسار
- تشبيك الأحرف

الترميز

ترميز العربية



-متعدد اللغات

يشمل العربية والفرنسية
والإنجليزية



مسألة الترميز في اللغات الأخرى

■ غير متعلقة بالعربية فقط

غربي (ISO-8859-1)
غربي (ISO-8859-15)
غربي (IBM-850)
غربي (MacRoman)
غربي (Windows-1252)
سلتي (ISO-8859-14)
يوناني (ISO-8859-7)
يوناني (MacGreek)
يوناني (Windows-1253)
ايسلاندي (MacIcelandic)
نرويجي (ISO-8859-10)
جنوب أوروبي (ISO-8859-3)

غرب أوروبية
شرق أوروبية
شرق آسيوية
جنوب شرق و غرب آسيوية
شرق أوسطية
يونيكود

محدد من المستخدم
يونيكود (UTF-16)

اكتشف آليا
ترميزات أخرى
خصص القائمة...

عربي (Windows-1256)
عربي (ISO-8859-6)
يونيكود (UTF-8) ●
عربي (ISO-8859-1)
عربي (Windows-1252)

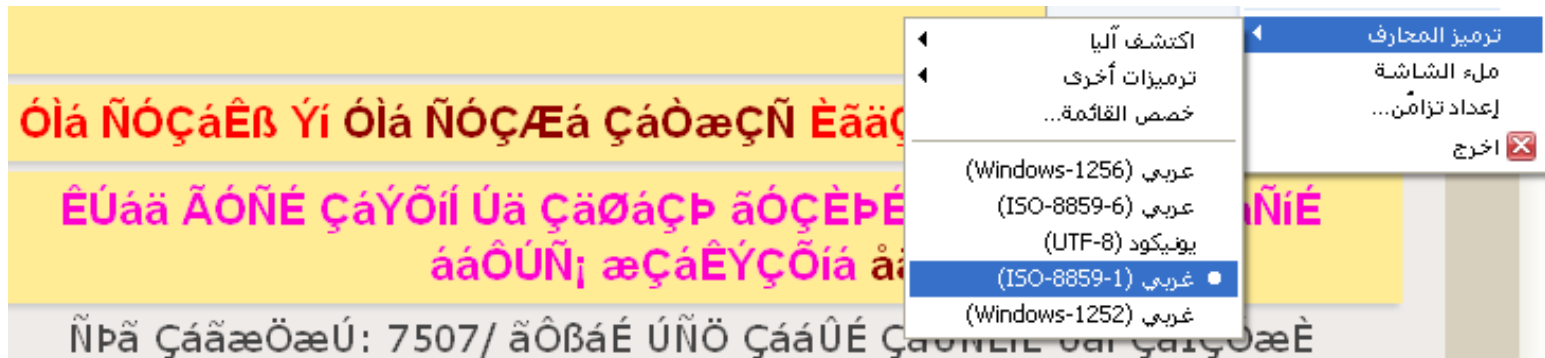
ترميز المخارف

ملء الشاشة
إعدادات تزامن...

اخرج

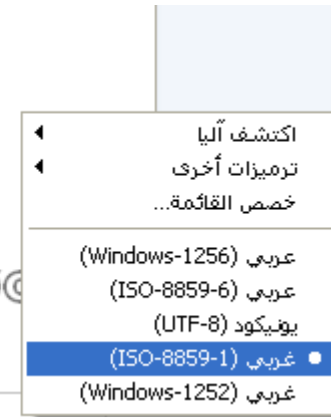
ترميز صفحات الويب

Windows 1256



UTF-8

Google
øšù,,ø¹ø±ø"ùšøø



Google
لشظ,,ظ1ط+ظ1ظظظ



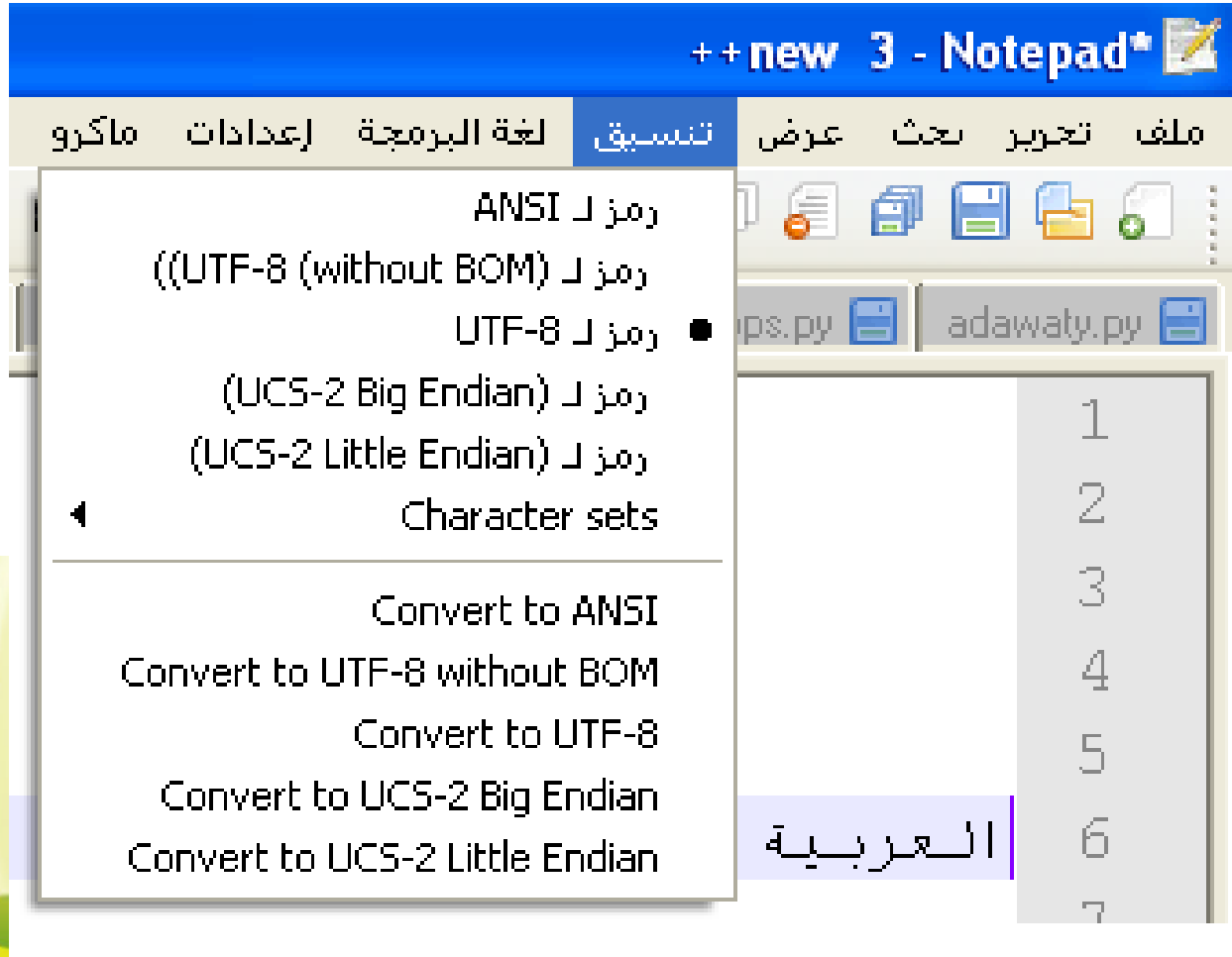
ترميز الملف HTML

■ لتحديد الترميز في ملف HTML نضع

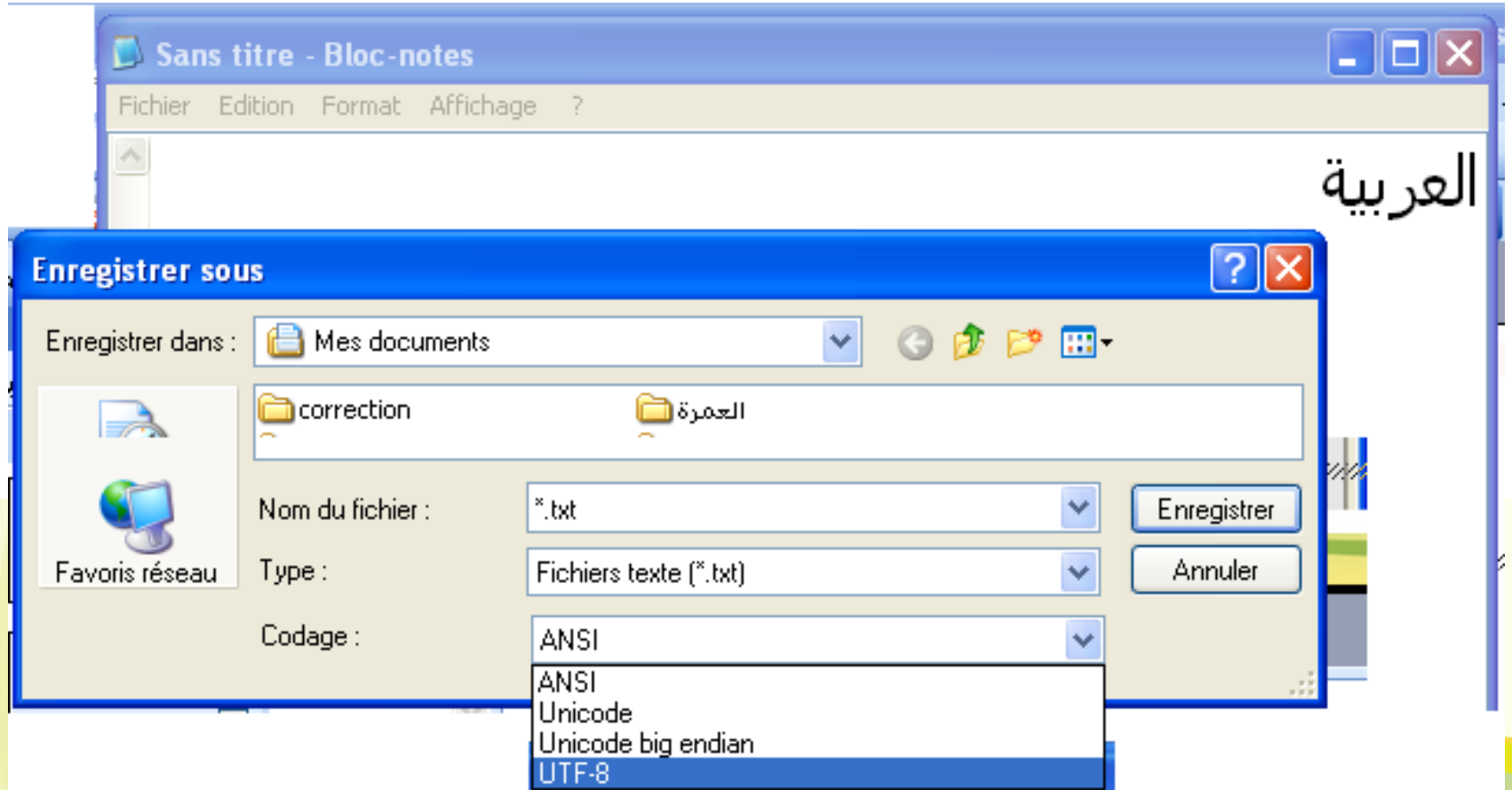
- `<head>`
- `<META http-equiv="Content-Type" content="text/html; charset=UTF-8">`
- `</head>`

ترميز الملفات

ترميز ملف



ترميز ملف

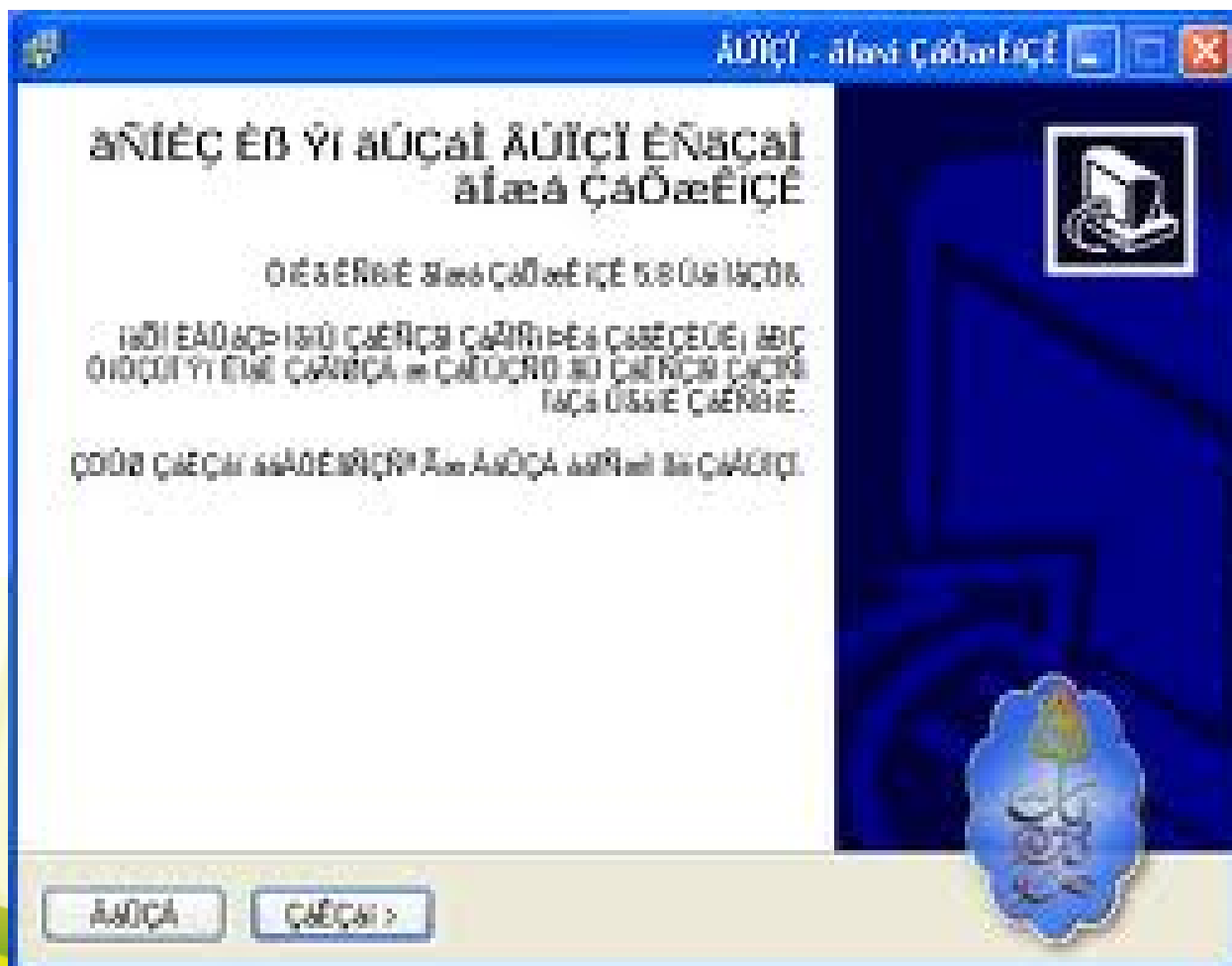


إعداد نظام التشغيل

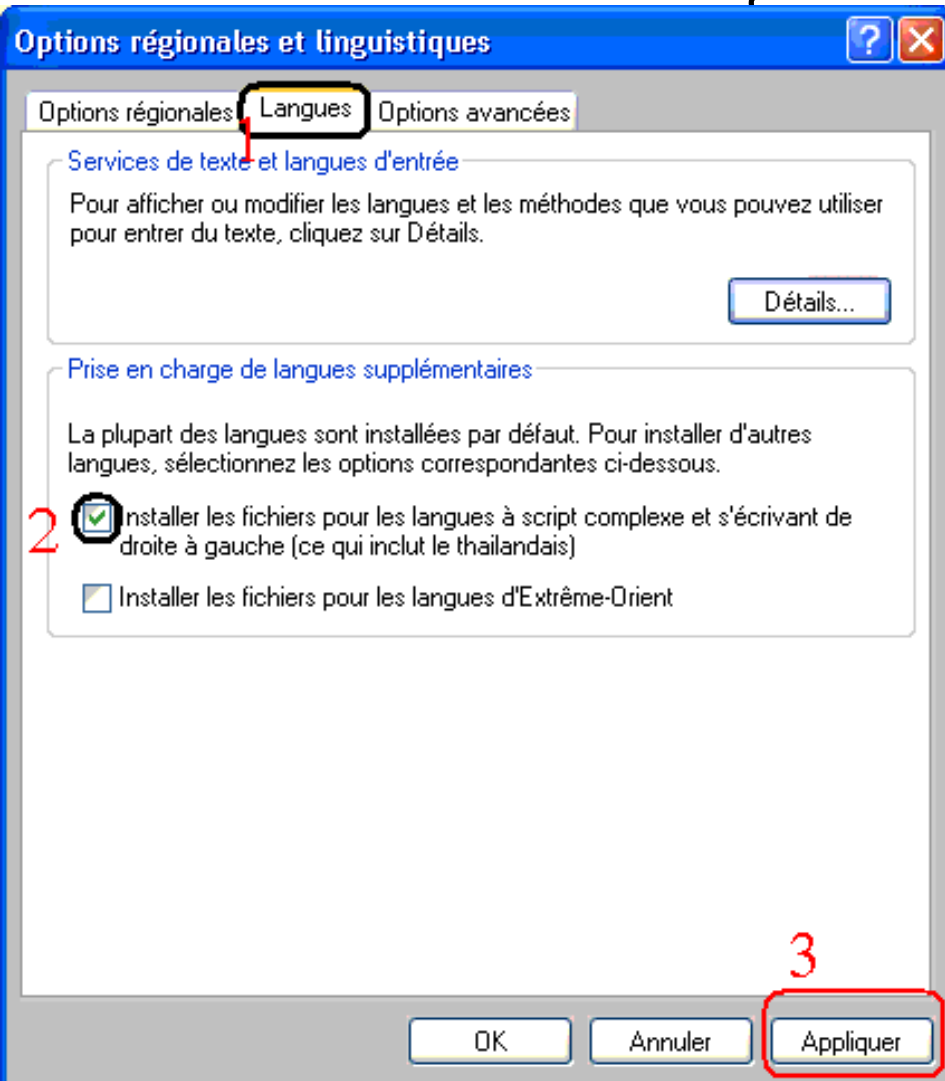
إعداد النظام

- إعداد النظام إذا كان لا يدعم العربية تلقائيا
 - ظهور رموز غريبة
 - تقطع الحروف

مشاكل عرض البرامج على وندوز



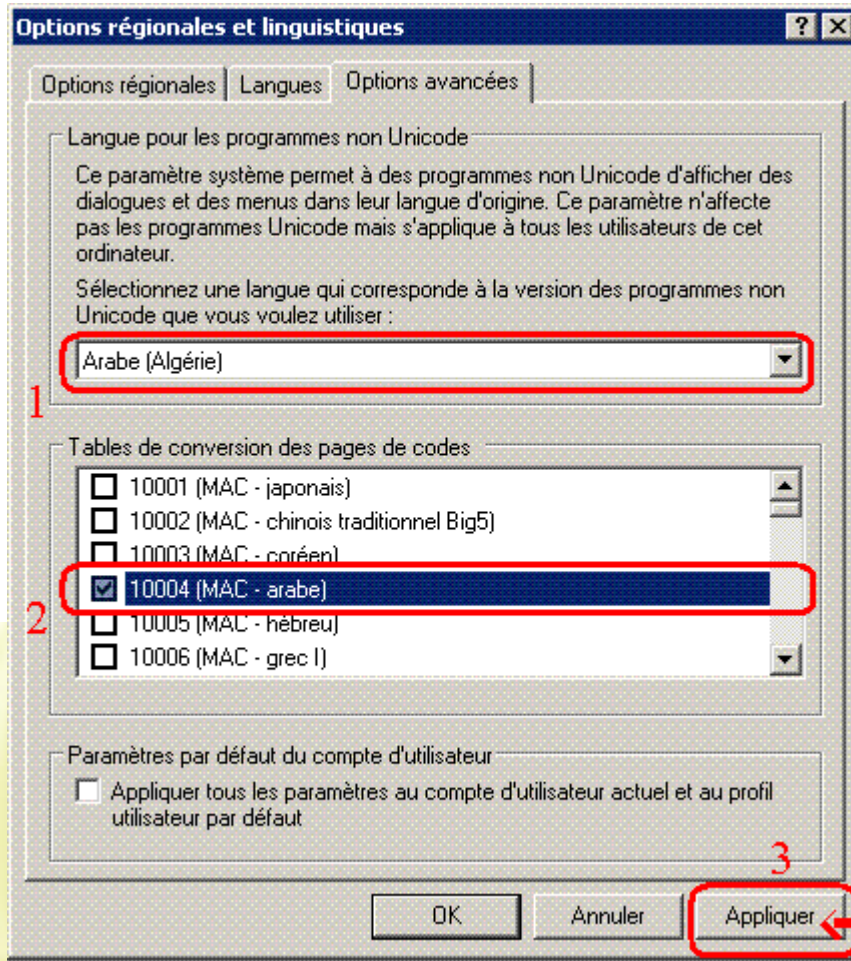
إعداد النظام



■ إعدادات إقليمية

■ تمكين النصوص مثل العربية

إعداد النظام



- للتعامل مع البرمجيات التي
- لا تستخدم يونيكود، تحديد
- الترميز التلقائي

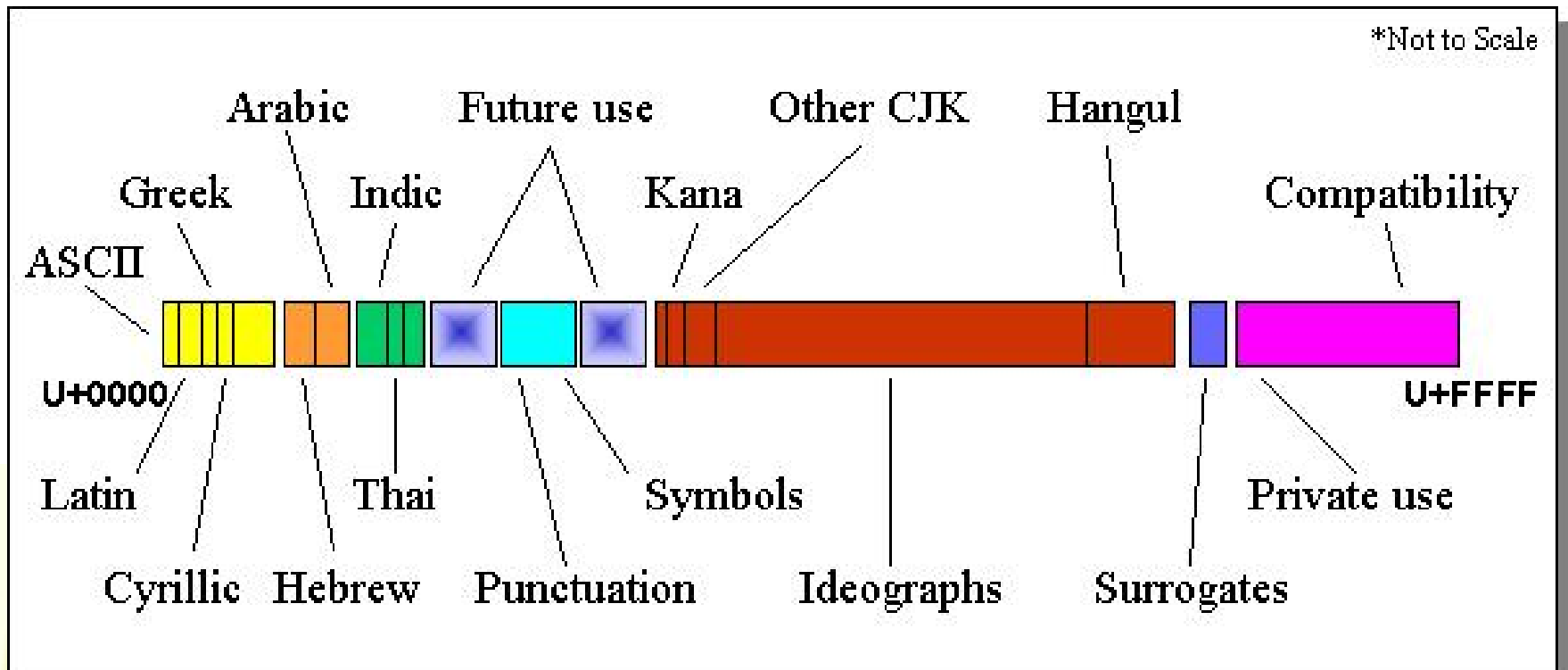
الترميز العالمي الموحد

Unicode

Unicode

- هي مجموعة رموز عالمية تستخدم لتعريف جميع الرموز والحروف المستخدمة في أغلب لغات العالم
- يستخدم من 1 إلى 4 بايت.
- قابل للإثراء
- Unicode.org

Unicode



تمثيل يونيكود

■ هناك ثلاثة أنواع رئيسية تستخدم حاليا لترميز Unicode:

■ UTF-8

وهو المفضل لدى مبرمجي الويب، حيث يستخدم 1 بايت إذا كانت الرموز موجودة في ترميز ASCII وتستخدم 2 إلى 4 بايت للرموز المعقدة.

■ UTF-16

هذا الترميز يستخدم إما 2 بايت لترميز إذا كانت الرموز موجودة في BMP (Basic Multilingual Plane) (و 4 بايت للرموز الغير موجودة).

■ **UTF-32** : يستخدم 4 بايت على الدوام.

تمثيل يونيكود

	A	⌘	好	𠄎
Code point	U+0041	U+05D0	U+597D	U+233B4
UTF-8	41	D7 90	E5 A5 BD	F0 A3 8E B4
UTF-16	41 00	D0 05	7D 59	D8 4C DF B4
UTF-32	41 00 00 00	D0 05 00 00	7D 59 00 00	B4 33 02 00

الحروف العربية في اليونيكود

- العربية الأساسية
- العلامات القرآنية
- العربية الموسعة
- الفارسية
- الأردو
- البشتو
- الكردية
- اللغات الإفريقية

■ التراكيب مثل لا، لم

العلامات القرآنية في اليونيكود

نص قرآني: سورة النمل من آية 25 إلى الآية 27:

أَلَّا يَسْجُدُوا لِلَّهِ الَّذِي يُخْرِجُ الْخَبَاءَ فِي السَّمَاوَاتِ وَالْأَرْضِ وَيَعْلَمُ مَا تُخْفُونَ وَمَا
تُعْلِنُونَ ﴿٢٥﴾ اللَّهُ لَا إِلَهَ إِلَّا هُوَ رَبُّ الْعَرْشِ الْعَظِيمِ ﴿٢٦﴾ قَالَ سَنَنْظُرُ أَصَدَقْتَ
أَمْ كُنْتَ مِنَ الْكَاذِبِينَ ﴿٢٧﴾

أحرف التحكم

- U+202A: LEFT-TO-RIGHT EMBEDDING
- U+202B: RIGHT-TO-LEFT EMBEDDING
- U+202D: LEFT-TO-RIGHT OVERRIDE
- U+202E: RIGHT-TO-LEFT OVERRIDE
- U+202C: POP DIRECTIONAL FORMATTING
- U+200E: LEFT-TO-RIGHT MARK
- U+200F: RIGHT-TO-LEFT MARK

التعامل مع التراميز

Python

■ يتعامل بيثون مع اليونيكود بطريقة شفافة

- `#!/usr/bin/python`
 - `# -*- coding=utf-8 -*-`
 - `string=u"العربية"`
 - `print len(string)`
 - `print string[3]`
-
- `>>7`
 - `>>ر`

PHP

- **Php لا يتعامل بطريقة شفافة مع اليونيكود، لذا نحتاج لبعض الأدوات المكتبات لتعامل معها، أو باعتبارها سلسلة بايتات**
- **الاستعانة ببعض المكتبات مثل mbstring**
- **مكتبة ar-php**

Ar-PHP.org



■ مزايا مشروع PHP واللغة العربية

- • التلخيص الآلي للنص العربي
- البحث بدلالة ساق الكلمة
- معالجة إظهار الحروف العربية المتصلة
- عرض التواريخ بالعربية أو الهجرية
- الترجمة اللفظية للكلمات
- التفقيط: كتابة الأرقام العربية



MySQL

■ إضافة استعلام MySQL هذا كي ينفذ مباشرة بعد الاتصال بقواعد البيانات:

- `set character_set_server='utf8';`
- `set names 'utf8';`

MySQL

■ أو بلغة php

- `<?php`
- `mysql_connect($db_hostname, $db_username, $db_password);`
- `mysql_query("set character_set_server='utf8'");`
- `mysql_query("set names 'utf8'");`
- *// The rest of you php code!*
- `?>`

العرض

العرض

ةيبرعلا

العربية

■ ثنائية الاتجاه

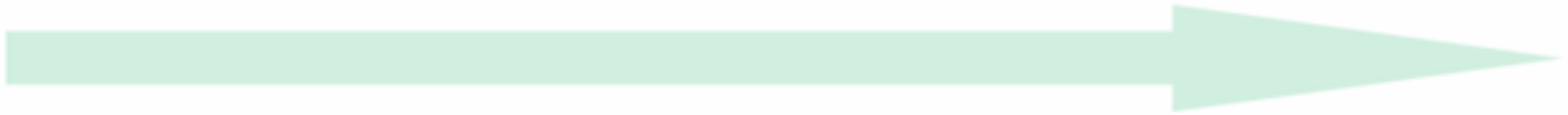
■ تشبيك الحروف

العربية

ثنائية الاتجاه

الأمير Charles والأميرة Camilla

Camilla وري مأل Charles وري مأل



الاتجاهات

- الصينية
- اليابانية
- الكورية

當世界需要溝通時，請用統一碼（Unicode）你現在就應報名將在1997年3月10至12日於德國美姿城（Mainz）召開的第十屆國際統一碼研討會。本次研討會將邀請多位業界專家研討關於全球網際網路及統一碼發展、國際化及本土化、支援統一碼的作業系統及應用程式、字型、文字排版、電腦多國語文化等多項課題。

الاتجاه HTML

■ استعمال محدد الاتجاه

- `dir = LTR | RTL`
 - LTR: Left-to-right text or table.
 - RTL: Right-to-left text or table.
- `<p dir='rtl'>`
- `<html dir='rtl'>`

CSS

- `div`
`{`
`direction:rtl;`
`}`
- Direction : rtl, ltr , inherit

أحرف التحكم

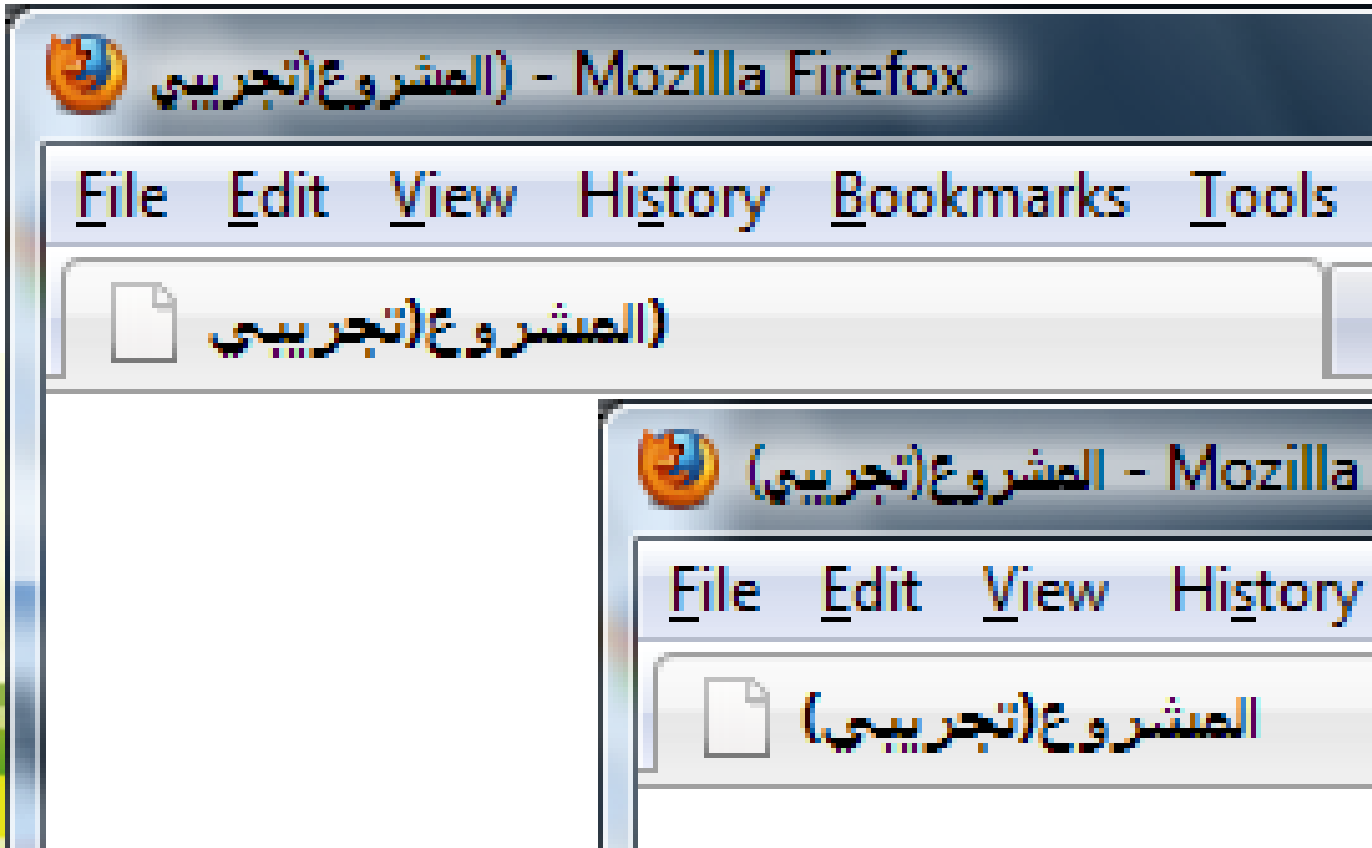
- U+202A: LEFT-TO-RIGHT EMBEDDING
- U+202B: RIGHT-TO-LEFT EMBEDDING
- U+202D: LEFT-TO-RIGHT OVERRIDE
- U+202E: RIGHT-TO-LEFT OVERRIDE
- U+202C: POP DIRECTIONAL FORMATTING
- U+200E: LEFT-TO-RIGHT MARK
- U+200F: RIGHT-TO-LEFT MARK

مثال

RLE

PDF

- `<title>‫المشروع(تجريبي)‬</title>`



العرض

ةيبرعلا

العربية

■ ثنائية الاتجاه

■ تشبيك الحروف

العربية

تشبيك الحروف

- اختيار شكل الحرف المناسب حسب الد
- يجرى عادة على مستوى الخط

لحم

لحم

تشابك الحروف في الهندية

ङ् + क → ङक

البرامج التي لا تدعم العربية

- بعض البرامج لا تستخدم إعدادات النظام في التعامل مع الحروف

البرامج التي لا تدعم العربية

■ استعمال برامج وسيطة



■ <http://tadween.sourceforge.net/>

برمجيا

■ لتطوير بعض التطبيقات التي لا تستخدم ما يوفر نظام التشغيل، أي لها محرك المعالجة الرسومية الخاص مثل:

□ برامج التصميم والرسم

□ الألعاب

■ نستعين بمكتبات Bidi

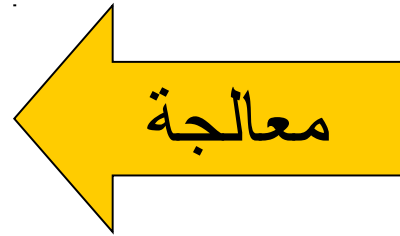
■ GNU Fribidi : fribidi.org

■ Python friBidi : pyfribidi.sf.net

خوارزمية BiDi

في الذاكرة

0	1	2	3	4	5	6
ا	ل	ع	ر	ب	ي	ة



في العرض

العربية

مثال لاستعمال pyfribidi



- لعبة كلمات عربية، اجمع الحروف لتكوين الكلمات المطلوبة على الشاشة

<http://arabicwordslide.sf.net>

pyfribidi

- `pyfribidi import log2vis`
- `self.image = self.font.render(log2vis(_SCORE+" %d" % self.score), 1, (255,255,255))`
- `Bidi(s)`
- `Shape(s)`

أحرف التحكم في التشبيك

■ كتاب التاريخ الهجري : 1433 هـ

□ هـ + - = هـ

□ هـ + = ZWJ هـ

■ Zero width joiner

■ U+200D

أحرف التحكم في التشبيك

■ ويك لولا الصبر ما كذبت ملأت الكيس تبرا

أحرف التحكم في التشبيك

■ المدرسة الوطنية العليا للإعلام الآلي

□ م.و.ع.إ.آ
م.و.ع.إ.آ

■ Zero width non joiner

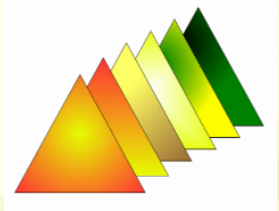
مشاريع

مشاريعنا



AraCorpus

مسكوك



تاشفين



PyArabic



غيوم

أدوات

نقطويه

غطاوي



آية سبل



مشكال

للاتصال

■ مدونتي <http://blog.tahadz.com>

للاتصال

■ Taha.zerrouki@gmail.com

معلومات وروابط المشاريع على مدونتي