بسم الله الرحمن الرحيم

www.nu.edu.pk

www.crulp.org

SSML for Urdu Speech Synthesis

Sarmad Hussain

Professor and Head
Center for Research in Urdu Language Processing
National University of Computer and Emerging Sciences, Lahore

Background

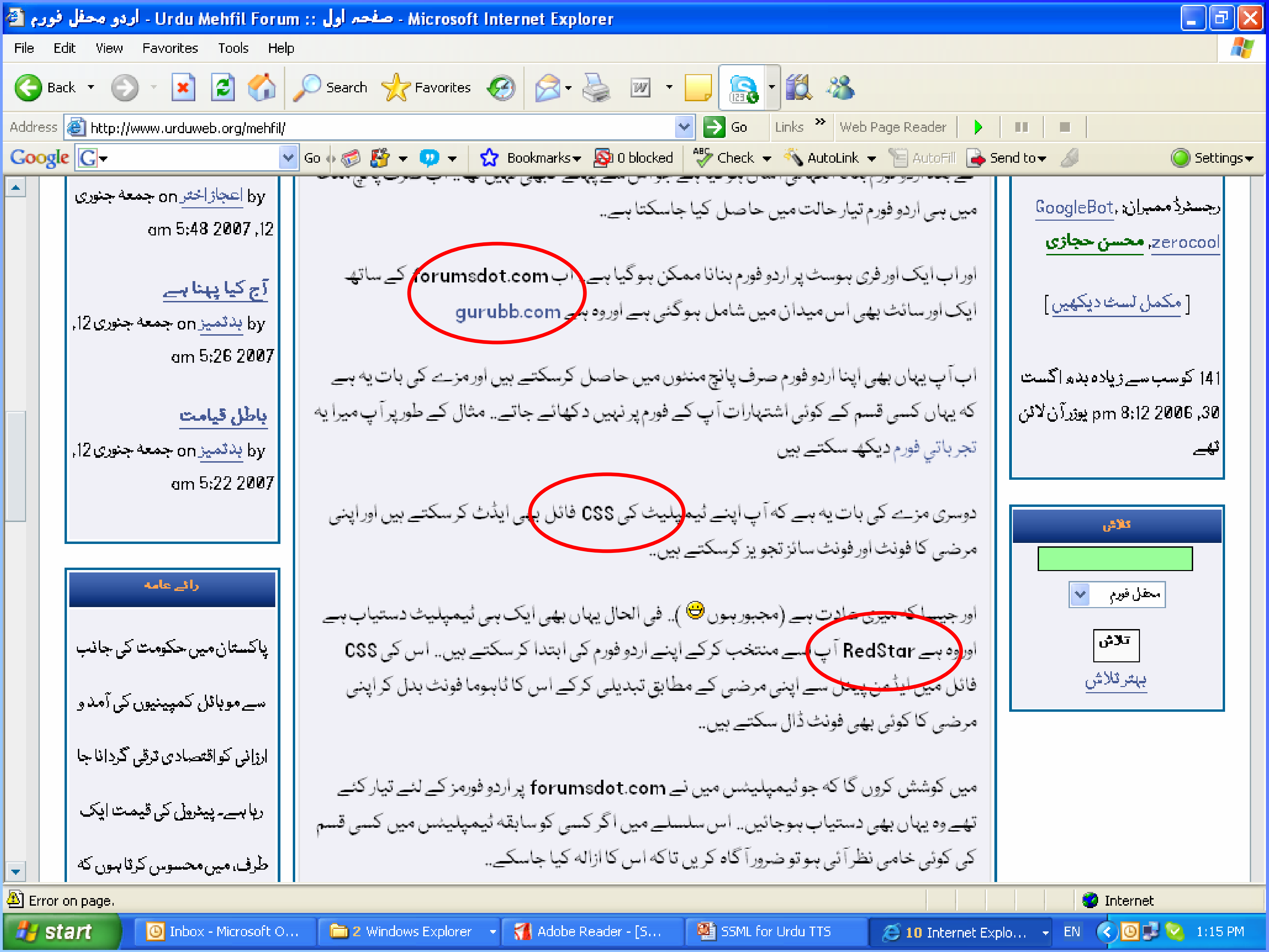
- Comments based on experience in developing a concatenative Urdu TTS system
  - inputs Urdu Unicode or UZT text and outputs Urdu speech
  - performs textual analysis, normalization and phonological processing to convert the text to speech
  - uses 5000+ diphone database
  - models duration and intonation based on learning through tagged corpus
  - complete framework developed by CRULP through funding by Ministry of IT, Govt. of Pakistan

Using SSML for Urdu Speech Synthesis

- SSM L standards works well for Urdu
- SSML needs to be enhanced to handle some Urdu specific problems
  - Multilingual text
  - Digits
  - Date formats
  - Diacritics
  - Word Segmentation

Multilingual Text

- Need to identify one of the three strategies (for different contexts)
  - Switch to English TTS
  - Transliterate and Read
  - Spell-out


اردو محفل فورم - Urdu Mehfil Forum :: صفحہ اول - Microsoft Internet Explorer
http://www.urduweb.org/mehfil/
by اعجاز اختر on جمعہ جنوری 12, 2007 5:48 am
آج کیا پہنا ہے
by بدتمیز on جمعہ جنوری 12, 2007 5:26 am
باطل قیامت
by بدتمیز on جمعہ جنوری 12, 2007 5:22 am
رائے عامہ
پاکستان میں حکومت کی جانب سے موبائل کمپنیوں کی آمد و ارزانی کو اقتصادی ترقی گردانا جا رہا ہے۔ پیٹرول کی قیمت ایک طرف میں محسوس کرتا ہوں کہ
میں ہی اردو فورم تیار حالت میں حاصل کیا جاسکتا ہے۔۔
اور اب ایک اور فری ہوسٹ پر اردو فورم بنانا ممکن ہوگیا ہے۔ اب forumsdot.com کے ساتھ ایک اور سائٹ بھی اس میدان میں شامل ہوگئی ہے اور وہ ہے gurubb.com
اب آپ یہاں بھی اپنا اردو فورم صرف پانچ منٹوں میں حاصل کرسکتے ہیں اور مزے کی بات یہ ہے کہ یہاں کسی قسم کے کوئی اشتہارات آپ کے فورم پر نہیں دکھائے جاتے۔۔ مثال کے طور پر آپ میرا یہ تجرباتي فورم دیکھ سکتے ہیں
دوسری مزے کی بات یہ ہے کہ آپ اپنے ٹیمپلیٹ کی CSS فائل بھی ایڈٹ کرسکتے ہیں اور اپنی مرضی کا فونٹ اور فونٹ سائز تجویز کرسکتے ہیں۔۔
اور جیسا کہ میری عادت ہے (مجبور ہوں 😃)۔۔ فی الحال یہاں بھی ایک ہی ٹیمپلیٹ دستیاب ہے اور وہ ہے RedStar آپ اسے منتخب کرکے اپنے اردو فورم کی ابتدا کرسکتے ہیں۔۔ اس کی CSS فائل میں ایڈمن پینل سے اپنی مرضی کے مطابق تبدیلی کرکے اس کا تاہوما فونٹ بدل کر اپنی مرضی کا کوئی بھی فونٹ ڈال سکتے ہیں۔۔
میں کوشش کروں گا کہ جو ٹیمپلیٹس میں نے forumsdot.com پر اردو فورمز کے لئے تیار کئے تھے وہ یہاں بھی دستیاب ہوجائیں۔۔ اس سلسلے میں اگر کسی کو سابقہ ٹیمپلیٹس میں کسی قسم کی کوئی خامی نظر آئی ہو تو ضرور آگاہ کریں تاکہ اس کا ازالہ کیا جاسکے۔۔
رجسٹرڈ ممبران: GoogleBot, zerocool, محسن حجازی
[مکمل لسٹ دیکھیں]
141 کو سب سے زیادہ بدھ اگست 30, 2006 8:12 pm یوزر آن لائن تھے
تلاش
محفل فورم
تلاش
بہتر تلاش
Error on page.
Internet
start
1:15 PM

أردو محفل - Urdu Mehfil Forum :: - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Address http://www.urduweb.org/mehfil/news_index.php?topic_id=968

أردو محفل

محفلِ دوستاں ، بزمِ یاراں

کیلنڈر آپ کے سوالات تلاش فہرستِ ارکان ارکان کے گروپ

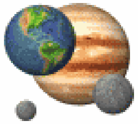
شمولیت اختیار کریں لاگ اِن

موجودہ وقت جمعہ جنوری 12, 2007 8:20 am ہے آپ کی آخری آمد سے بعد کے خطوط

تمام اوقات 0 + UTC جن خطوط کے جوابات نہیں دئیے گئے

نیوز مضامین

| کٹیگریز | محفوظات



مریخ پر بھیجی گئی گاڑی کے دو سال مکمل

2006-03-01 05:45:45 pm by عتیق الرحمن | (Comments (0

4 جنوری 2006 کو مریخ پر بھیجی گئی روبوٹ گاڑی (Robot Rover) کے دو سال مکمل ہو جائیں گے ----
4 جنوری 2004 بوقت 0435GMT کو امریکی خلائی ادارے ناسا نے مریخ پر سپرٹ (Spirit) نامی روبوٹ گاڑی اتاری تھی اور پھر 25 جنوری 2004 کو ایک اور روبوٹ گاڑی Opportunity نے مریخ کے قدم چومے ---- ناسا کا خیال تھا یہ گاڑیاں وہاں 3 ماہ تک کارگر رہیں گی ---- مگر اب بھی وہ اس سرخ سیارے پر مٹر گشت کر رہیں ہیں ۔ ان گاڑیوں کو مریخ پر اتارنا ناسا کی بڑی کامیابیوں میں سے ایک کامیابی ہے ---- اور پھر جن مقاصد کے لیئے ان گاڑیوں کو خلا میں بھیجا گیا تھا وہ بھی پورے ہوئے اور ان گاڑیوں کی وجہ سے ہم یہ معلوم کر سکے کہ مریخ پر کسی دور میں پانی بہتا رہا ہے اور اس نظریے کو تقویت ملی کہ مریخ کے قطب میں جما ہوا پانی موجود ہے ۔

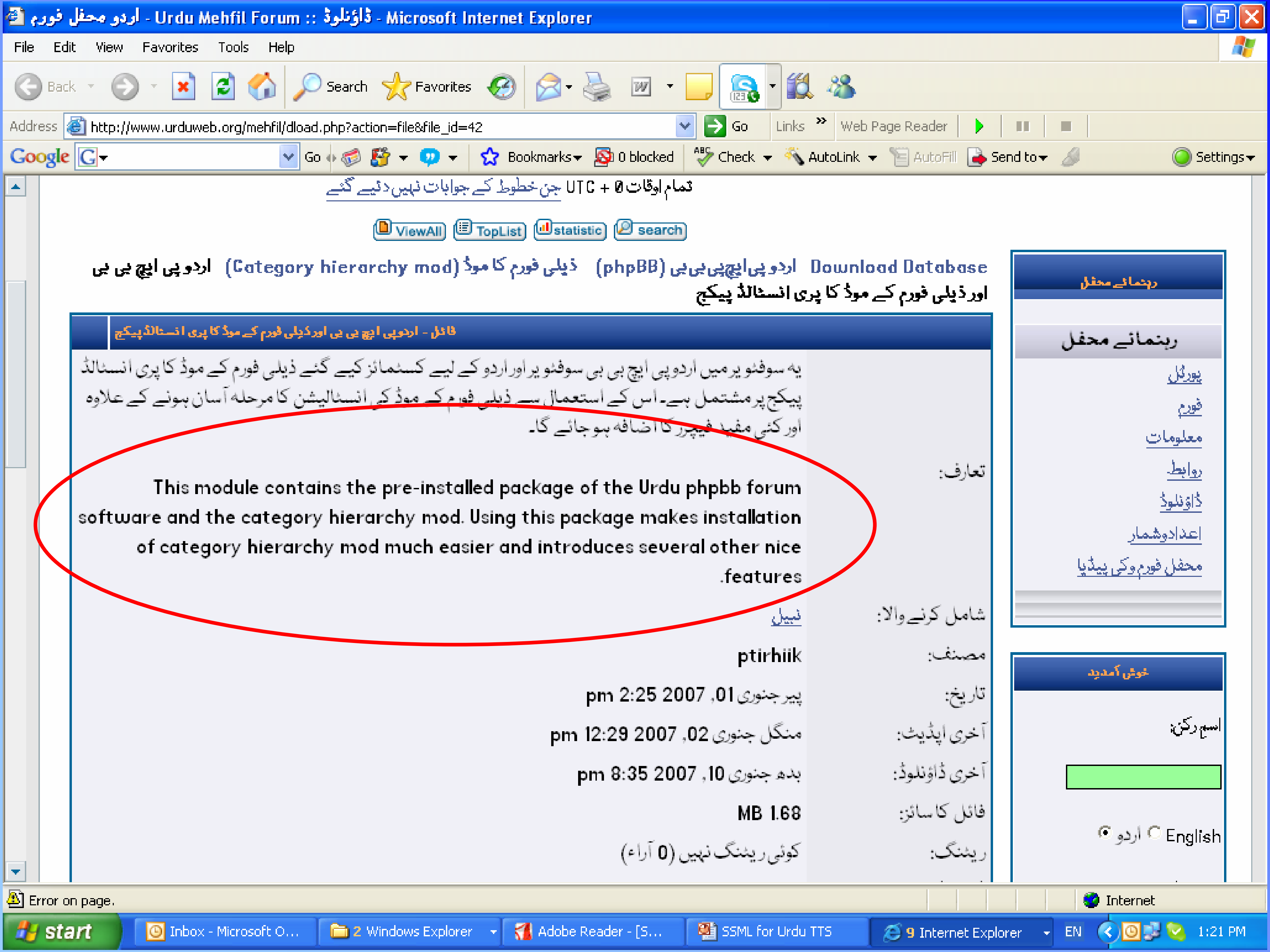
رہنمائے محفل

رہنمائے محفل

- پورٹل
- فورم
- معلومات
- روابط
- ڈاؤنلوڈ
- اعداد و شمار
- محفل فورم وکی پیڈیا

Error on page. Internet

start | Inbox - Microsoft O... | 2 Windows Explorer | Adobe Reader - [S... | SSML for Urdu TTS | 10 Internet Explo... | EN 1:18 PM


اردو محفل فورم - Urdu Mehfil Forum :: ڈاؤنلوڈ - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Address http://www.urduweb.org/mehfil/dload.php?action=file&file_id=42
تمام اوقات UTC + 0 جن خطوط کے جوابات نہیں دئیے گئے
ViewAll TopList statistic search
Download Database اردو پی ایچ پی بی بی (phpBB) ذیلی فورم کا موڈ (Category hierarchy mod) اردو پی ایچ پی بی بی اور ذیلی فورم کے موڈ کا پری انسٹالڈ پیکج
فائل - اردو پی ایچ پی بی بی اور ذیلی فورم کے موڈ کا پری انسٹالڈ پیکج
تعارف:
یہ سوفٹویر میں اردو پی ایچ پی بی سوفٹویر اور اردو کے لیے کسٹمائز کیے گئے ذیلی فورم کے موڈ کا پری انسٹالڈ پیکج پر مشتمل ہے۔ اس کے استعمال سے ذیلی فورم کے موڈ کی انسٹالیشن کا مرحلہ آسان ہونے کے علاوہ اور کئی مفید فیچرز کا اضافہ ہو جائے گا۔
This module contains the pre-installed package of the Urdu phpbb forum software and the category hierarchy mod. Using this package makes installation of category hierarchy mod much easier and introduces several other nice features.
شامل کرنے والا: نبیل
مصنف: ptirhiik
تاریخ: پیر جنوری 01, 2007 2:25 pm
آخری اپڈیٹ: منگل جنوری 02, 2007 12:29 pm
آخری ڈاؤنلوڈ: بدھ جنوری 10, 2007 8:35 pm
فائل کا سائز: 1.68 MB
ریٹنگ: کوئی ریٹنگ نہیں (0 آراء)
رہنمائے محفل
پورٹل
فورم
معلومات
روابط
ڈاؤنلوڈ
اعداد و شمار
محفل فورم وکی پیڈیا
خوش آمدید
اسم رکن:
اردو English
Error on page.
Internet
start
1:21 PM

اِنڈونیشیا :سیلاب اور مٹی کے تودے گرنے سے کم از کم 18 افراد ہلاک - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Address http://209.41.165.188/urdu/update_details.asp?nid=3015

Enter Search Keywords

Advance Search

جنگ اردو Searchable

Download Font | Urdu Keyboard | On-Screen Keyboard | Back Issues | Feedback   Friday, January 12, 2007, Zil-Haj 21, 1427 A.H

تازہ ترین

انڈونیشیا :سیلاب اور مٹی کے تودے گرنے سے کم از کم 18 افراد ہلاک

Updated at 08:00 PST

جکارتہ۔۔۔۔انڈونیشیا میں سیلاب اور مٹی کے تودے گرنے کے باعث کم از کم اٹھارہ افراد ہلاک ہوگئے ۔جبکہ درجنوں افراد لاپتہ ہیں ۔جکارتہ میں حکام کے مطابق واقعہ انڈونیشیا کے جزیرے Sulawesiمیں جمعے کی صبح اس وقت پیش آیا جب سیلاب اور مٹی کے تودے گرنے کے نتیجے میں متعدد افراد ہلاک ہوگئے ۔جبکہ درجنوں افراد تاحال لاپتہ ہیں ۔اسپتال ذرائع نے کم از کم اٹھارہ افراد کے ہلاک ہونے کی تصدیق کی ہے ۔سیلاب اور مٹی کے تودے گرنے کے بعد سیکڑوں افراد گھروں کو چھوڑ کر اسکول کی عمارتوں میں پناہ لینے پر مجبور ہوگئے ہیں ۔

Print Version

القاعدہ رہنماؤں کی پاکستان میں موجودگی کا الزام مسترد	Updated at 13:20 PST
امریکی فوجی کو عراقی قیدیوں کے قتل میں 18 سال قید کی سزا	Updated at 13:00 PST
پی ایس او کی نجکاری مارچ تک مکمل کرلی جائے گی،زاہد حامد	Updated at 12:40 PST
کراچی اسٹاک :ٹریڈنگ کے دوران انڈیکس میں28پوائنٹس کا اضافہ	Updated at 12:20 PST
عراق:چوبیس گھنٹوں کے دوران 37 لاشیں ملیں،حکام	Updated at 12:05 PST

بہتر فونٹ کے لئے یہاں کلک کریں

Download & Installation Instructions

- آج کا اخبار
- تازہ ترین
- اہم خبریں
- ملک بھر سے
- شہر قائد / شہر کی آواز
- سندھ بھر سے
- دنیا بھر سے
- امریکا سے
- یورپ سے
- ادارتی صفحہ
- اسپورٹس
- بزنس

Done   Internet

start   Inbox - Microsoft O...   2 Windows Explorer   Adobe Reader - [S...   SSML for Urdu TTS   7 Internet Explorer   EN   1:22 PM

Digits

- Should be able to read out Urdu digits
  - Issues related to normalization?
- Two strategies for English digits (to be defined by the website)
  - Read in English
  - Read in Urdu

Date Formats

- Cover other date formats in Urdu
  - A complete list can be provided
- Cover two calendars
  - Lunar based islamic calendar
  - Regular calendar

Links
Web Page Reader
blocked
Check
AutoLink
AutoFill
Send to
Settings
اردو
Searchable
جنگ
Feedback
Friday, January 12, 2007, Zill-Haj 21, 1427 A.H
صومالیہ: 2دن کی امریکی بمباری
BBCUrdu.com - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back
Search
Favorites
Address
http://www.bbc.co.uk/urdu/regional/story/2007/01/070110_euplan_industry_revolution_sz.shtml
Go
Links
Google
Bookmarks
0 blocked
Check
AutoLink
BBC
Home
News
Sport
Radio
TV
Weather
Languages
BBC URDU.com
Wednesday, 10 January, 2007, 15:00 GMT 20:00 PST وقتِ اشاعت:
صفحہِ اول
پاکستان
انڈیا
یورپ : انرجی کی نئی حکمت عملی

Diacritics

- Optionally, incompletely and (sometimes) incorrectly given
- تقریباً (almost), اعلٰی (high (in quality or rank))
- بَن بِن بُن بن (ambiguous, knit (v), son of (particle), make (v))

Diacritics

- Lexical look-up
  - Should be ignored, and look up base form from the lexicon
  - MUST match given diacritics (even if incomplete)
    - Normally given where diacritic can cause ambiguity
- If lexical look-up fails, use a pronunciation guessing engine?
    - Use default mechanism/ No engine
    - Use an engine, at the given path: need to decide on
      - diacritics or pronunciation?
        - » IPA, Custom Scheme, others?
      - Accuracy vs. performance?
        - » POS tagger
        - » WSD

Word Segmentation

- Urdu does not have the concept of SPACE
- SPACE used by typists to get the visual output
  - Spaces within words
    - یونی ورسٹی
  - No spaces between words, esp. after non-joiners, as words visually OK
    - اآدڈذرڑزژو
- Should it point to word segmentation engine to enable correct output?
  - Use default mechanism/ No engine
  - Use an engine, at the given path: need to decide on
    - Accuracy vs. performance?

بسم الله الرحمن الرحيم

www.nu.edu.pk

www.crulp.org

SSML for Urdu Speech Synthesis

Sarmad Hussain

Professor and Head
Center for Research in Urdu Language Processing
National University of Computer and Emerging Sciences, Lahore