Java Multilingual Elementary Tool

Komminist Sisai(04305201)

November 28, 2004

Outline

- Introduction
- Existing Approach
- Designing
- Creating New Language
- Graph and Picture
- References

Introduction

- **Multilingual system:**
    - refer to computer programs which permit user interaction with the computer in one or more languages
- **A Java multilingual elementary tool** (JMET) is computer based **language teaching tool.**
- Main objectiveis to:
    - **teach languages** for elementary students

Benefit of Learning others Language

Learning some other language has a number of advantageous

1. **Personal Benefits:**
   - provides expanded access to people and resources,
   - ability to communicate with more people,
   - introducing students to alternative ways of expressing themselves
   - opening up additional job opportunities, and many more...
2. **Cognitive Benefits:**
   - more creative and better at solving complex problems
3. **Academic Benefits**
4. **Societal Benefits**

Existing System

- availability of different language in the world
- needs to represent them in computer system so that standardization is maintained
- **There are a number of existing system:**
  - **but all share the use of Unicode**
- for example IITM develope multilingual system which:
  - permit data entry and display in Indian Languages
- so that we needs to study Unicode

What is Unicode? I

What is Unicode?

- Representing text-format data in computers is a matter of defining:
    - set of characters and assigning each of them a number,
- Unicode provides a unique number for every character,
    - no matter what the platform,
    - no matter what the program,
    - no matter what the language.
- its a universal encoded character set

What is Unicode? II

Unicode Version

- early versions of Unicode
  - used 16-bit numbers,
  - which allowed for encoding 65,536 characters
- with the inclusion of more characters, we need to extend Unicode
- Several other encoding schemes were added

Unicode Version continued...

- at Unicode version 2.0,
  - there were 38,885 assigned characters
- at version 3.0
  - there were 49,194 assigned characters
- at version 3.2
  - there were 95,156 assigned characters
- at version 4.0
  - there are 96,382 assigned characters

Unicode Encodings

There are several character encoding defined by **ISO 10646** for Unicode:

- **UTF-8** : Unicode Transfer Formate
  - 8-bit Unicode Transformation Format,
  - represents each character in the range U+0000 through U+007F as a single octet
  - it use a group of bytes to represent Unicode standard,
  - it uses 1 to 4 bytes per character, depending on the Unicode symbol

  example: ASCII

Unicode encoding continue....

- **UTF-16** : Unicode Transfer Formate
  - uses value between 0-65535
  - represented as exactly two bytes,
  - for example, 0 for 0, FFFD for FFFD.
  - What if character above FFFF / 65535 ?
    - represents as a surrogate pair of code values from the range D800-DFFF

UTF-16 Examples			
Code Point	Character	UTF-16 value(s)	glyph
122(hex 7A)	small z(Latin)	007A	z
27700 (hex 6C34)	water (Chinese)	6C34	水
119070 (hex 1D11E)	musical G clef	D834 DD1E	01D 11E

Unicode encoding continue....

- **UTF-32** : Unicode Transfer Formate
  - between 0 and FFFFFFFF for each character,
  - represented as exactly four bytes,

Range of Unicode

Range of Unicode

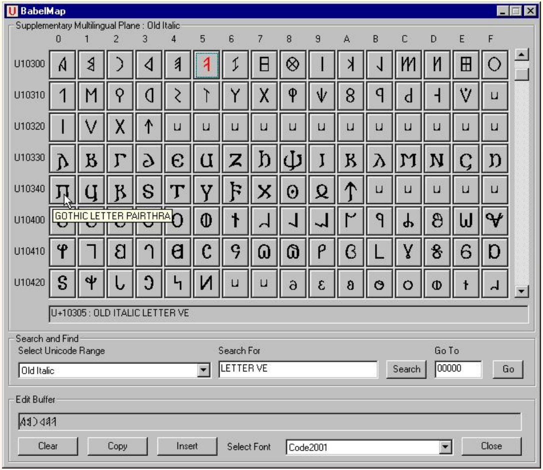
- Each language has its own predefined Unicode character range
- based on this range we can access them in any machine that support Unicode
- **Examples of Unicode range:**
  - **U+1200 U+137F (46084991): Ethiopic character range**
  - **U+0600 U+06FF (15361791): Arabic character range**
  - **U+0370 U+03FF (8801023) : Greek**
  - **etc...**

Unicode and Multilingual File Conversion,Font Utilities and Editors

For Windows Computers examples

- Babel-Map
  - its Unicode character map from Windows 95 onwards
  - includes Unicode 3.2 version
  - can displays characters from the supplementary panels even under windows 95
- **Unicode Font Viewer**
  - its freeware
  - runs under windows 9x, windows NT 4 and Windows 2000
  - allows you to display the Unicode 2.1 characters from any TrueType Unicode font.
  - an enlarged view of any character can be shown.

Unicode and Multilingual File Conversion, Font Utilities, Editor



Supplementary Multilingual Plane : Old Italic

	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
U10300	𐌀	𐌁	𐌂	𐌃	𐌄	𐌅	𐌆	𐌇	𐌈	𐌉	𐌊	𐌋	𐌌	𐌍	𐌎	𐌏
U10310	𐌐	𐌑	𐌒	𐌓	𐌔	𐌕	𐌖	𐌗	𐌘	𐌙	𐌚	𐌛	𐌜	𐌝	𐌞	u
U10320	𐌠	𐌡	𐌢	𐌣	u	u	u	u	u	u	u	u	u	u	u	u
U10330	𐌰	𐌱	𐌲	𐌳	𐌴	𐌵	𐌶	𐌷	𐌸	𐌹	𐌺	𐌻	𐌼	𐌽	𐌾	𐌿
U10340	𐍀	𐍁	𐍂	𐍃	𐍄	𐍅	𐍆	𐍇	𐍈	𐍉	𐍊	u	u	u	u	u
U10400	𐐀	𐐁	𐐂	𐐃	𐐄	𐐅	𐐆	𐐇	𐐈	𐐉	𐐊	𐐋	𐐌	𐐍	𐐎	𐐏
U10410	𐐐	𐐑	𐐒	𐐓	𐐔	𐐕	𐐖	𐐗	𐐘	𐐙	𐐚	𐐛	𐐜	𐐝	𐐞	𐐟
U10420	𐐠	𐐡	𐐢	𐐣	𐐤	𐐥	u	u	𐐨	𐐩	𐐪	𐐫	𐐬	𐐭	𐐮	𐐯

GOTHIC LETTER PAIRTHRA

U+10305 : OLD ITALIC LETTER VE

Search and Find

Select Unicode Range: Old Italic

Search For: LETTER VE

Go To: 00000

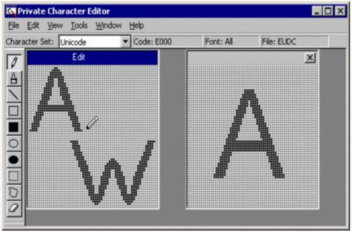
Edit Buffer

𐌀𐌁𐌂𐌃𐌄𐌅

Select Font: Code2001

For Windows Computers continued...

- **Private Character Editor**
  - allows you to create characters
  - and assign them to code points in the Private Use Area(U+E000 - U+F8FF)
  - you can either create a character from scratch, or
  - select an existing character from any installed font and modify it
  - the new characters can be used in any program that supports Unicode
  - its included in Win 2000 and Win Xp


Private Character Editor

Character Set: Unicode
Code: E000
Font: All
File: EUDC

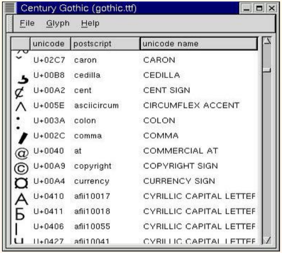
For Unix and Linux Computers examples

- **FontForge**
  - its formerly known as PfaEdit
  - is a freeware font editor for Unix and Linux
  - with it we can create and edit
    - TrueType,
    - OpenType,
    - bitmap (.bdf) and some Postscript fonts,
  - it can also convert between formats.

For Unix and Linux Computers continued...

- **GOTE**
  - GNOME OpenTyepe Editor
  - freeware TrueType font editor for GNOME environment under POSIX
  - still under development

Many more like: ttf2bdf, xfsft, XmBDFEd

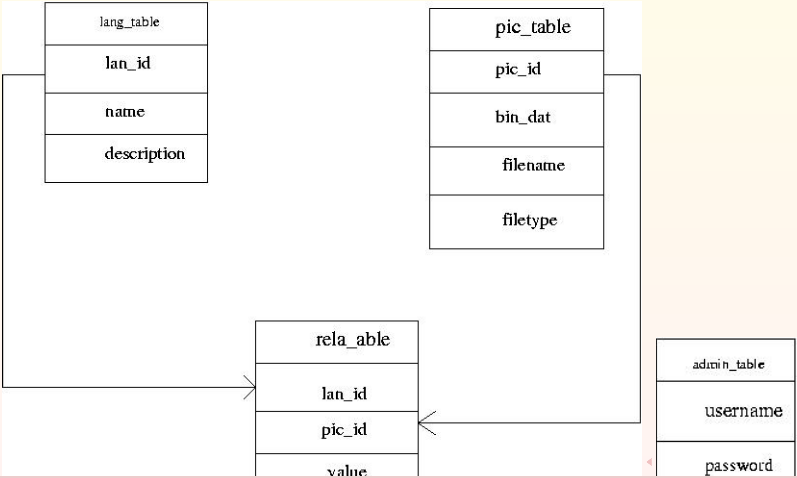


	unicode	postscript	unicode name
ˇ	U+02C7	caron	CARON
¸	U+00B8	cedilla	CEDILLA
¢	U+00A2	cent	CENT SIGN
^	U+005E	asciicircum	CIRCUMFLEX ACCENT
:	U+003A	colon	COLON
,	U+002C	comma	COMMA
@	U+0040	at	COMMERCIAL AT
©	U+00A9	copyright	COPYRIGHT SIGN
¤	U+00A4	currency	CURRENCY SIGN
А	U+0410	afii10017	CYRILLIC CAPITAL LETTEF
Б	U+0411	afii10018	CYRILLIC CAPITAL LETTEF
І	U+0406	afii10055	CYRILLIC CAPITAL LETTEF
Ч	U+0427	afii10041	CYRILLIC CAPITAL LETTEF

Our Approach

- Use the same concept of the existing approach,
  - **use of Unicode standards**
- additional feature added since the product focus on elementary students
  - instead of teaching language as usual, we use pictures

General Schema Diagram



language schema

table description

- *lan_id* : has the language id
- *name* : specifies the language name
- *description*: language description

how in Mysql

```
create table language(
lan_id int not null,
name char(20),
description char(20), primary key
(lan_id));
```

language schema continue...

Field	Type	Null	Key	Default	Extra
lan_id	int(11)		PRI	0	
name	char(20)	YES		NULL	
description	char(20)	YES		NULL	

How We Store Pictures In Database?

"developing database application requires more than just operation with textual or numeric data"

- for example, developing a multilingual based application
- frequently there is a need to display pictures along with text from a database
- **two existing models** -in reality not only two
  1. store the pictures in a folder and store the path to each one in a database
  2. store the entire file into a database, along with its file name
- each of them have its own ups and downs

Picture Schema

table description

- pic_id : has the picture id
- *description* : has picture description
- *bin_data* : picture data
- *filename* : file name of the picture- absolute path
- *filetype* : specifies the type of the picture as jif, jpg

how to create in Mysql

```
create table picture(
pic_id int not null,
description char(20),
bin_data blob,
filname char(20),
filetype char(10), primary key
(pic_id));
```

Picture Schema continue....

Field	Type	Null	Key	Default	Extra
pic_id	int(11)		PRI	0	
bin_data	blob	YES		NULL	
filename	char	YES		NULL	
filetype	char(10)	YES		NULL	

relation Schema

table description

- *lan_id* : references language id
- pic_id : references picture id
- *value*: corresponding language value

how to create in Mysql

```
create table muliling(
lan_id int not null,
pic_id int not null,
value varchar(10), primary key
(lan_id,pic_id));
```

relation Schema continue....

Field	Type	Null	Key	Default	Extra
lan_id	int(11)		PRI	0	
pic_id	iint(11)			0	
description	char(20)	PYES		PNULL	

Admin Schema

The schema structure of the admin relation look like

- *username* : the user name of the adiminstrator
- *password* : pass word
- *description*: description of the admin

how to create in Mysql

create table admin(
username char(15) not null,
password char(15) not null,
description char(20), primary key (usename,password));

adimin Schema continue....

Field	Type	Null	Key	Default	Extra
username	char(15)		PRI	0	
password	char(15)			0	
description	char(30)	YES		NULL	

User interfaces

- **user interface one**
  - displays a list of pictures
  - list of languages
  - navigation buttons
- **user interface two**
  - display the query result

User interface continued...

- **user interface three**
  - allows the admin to authenticate, so update is possible
- **user interface four**
  - allows the admin to create a new language - follows if authentication OK!

What is Required from Students

- **Students must be able to:**
  - read and write English language
  - know basic skills of computers (example how click, enter text)
- he/she must select picture and particular language

Sample result

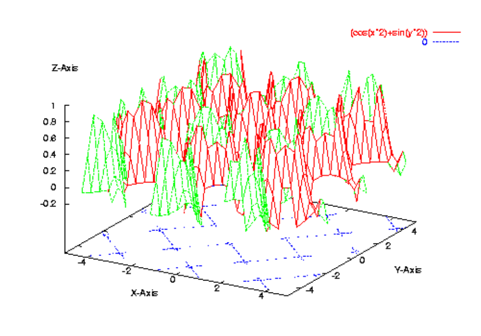
Input		Output
Picture	In English	In Amharic
	House	መለመር

Creating of new langauge

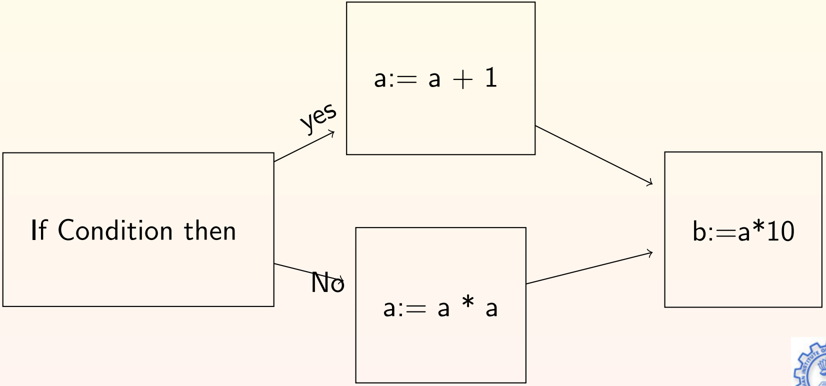
- **the admin do this job**
- but how?—- by drawing chracteres
- to draw there are different techniques
    - my approach is using bitmap drawing

Application of Gnuplot

function (cos(x*2)+sin(y*2))



pgf picture



References

1. http://www.alanwood.net/unicode/
2. http://www.unicode.org/
3. http://www.abyssiniacybergateway.net/fidel/fidel.html
4. http://acharya.iitm.ac.in/multi_sys/exist_codes.html
5. http://emeld.org/school/classroom/unicode/
6. http://shlimazl.nm.ru/eng/fonts_ttf.htm#nametab