UNIVERSITY OF MICHIGAN
Survey Research Center

Blaise Source Code Editing System

Presenter: Danilo Gutierrez
Co-author: Sheila Deskins
Health and Retirement Study (HRS)

The 11th International Blaise Conference
Annapolis, Maryland
September 2007

Presentation Overview

How Big is Big?

What Does a Source Editor Do?

The System and Updating a New Language

Current Use & Future Plans

Questions

How Big is Big?

HRS CAI Size

- Datamodel Source Code (.bla, .inc)

175,624	**Program lines**
61	Include files
518	Procedures
344	Blocks
10	Tables

HRS CAI Size

- **Fields**
  - **5,818 Fields**
  - 1,773 Auxfields
  - 1,691 Locals
  - 5,754 Parameters

HRS CAI Size

- **Type Definitions**
  - 8,962 USER-DEFINED
  - **1,390 ENUMERATED**

  - 2,366 STRING
  - 481 RANGE
  - 366 OPEN
  - 309 INTEGER
  - 126 ARRAY
  - 141 SET
  - 11 REAL

What does a “Source Editor” do?

System Core

- Parses source code files (.bla and .inc)
- Merges in update information
- Writes updated source code files

Parsing - example

Statement:

Q1 (One) "Are you ready to answer questions?“ : (y,n)

Parses into tokens of:

Q1
(One)
“Are you ready to answer questions?“
:
(
y
,
n
)

Definition of Token

A **token** is part of a program statement consisting of characters identified as meaningful syntax.

Merging

- There are two inputs in merging
  - The **user update** request information
  - The **parsed tokens** from original source code

User Request

User update request information:

fieldname: Q1

token type: descriptor

language: default

edit instruction: add (or delete)

update: **“This is the new label”**

Merge results - example

Source code in tokens:

Q1
(One)
“Are you ready to answer questions?”
“This is the new label”
:
(
y
,
n
)

Problem: the descriptor looks like question text!

How to fix the problem?

- Blaise Data Object (BDO) contains all **possible** parts of a Blaise syntax statement.

Blaise Syntax for **Fields**

Q [Q1, [...]] [(Tag)] [[Lid] "Text"] [...]
[/ [Lid] "Description"] [...] : T

Using the BDO

Syntax	BDO with update
Q	Q1
[Q1, [...]]	<blank>
[(Tag)]	(One)
[[Lid] "Text"]	[[Lid] "Are you ready to answer questions?"]
[...]	[...]
[/ [Lid] "Description"]	[**/** [Lid] **" This is the new label"**]
[...] : T	[...] : T

Edited Source Code

Q1 (One) “Are you ready to answer questions?”

/ “This is the new label” : (y , n)

Writing

- Writing is simpler when the database is already organized by the parsing and merging processes
- Need to write out whitespace, comments, file names, etc.
- Write Spanish language diacriticals
- Write the same number of files as were parsed

System Considerations

System Considerations

	Editing requires	.BMI (BCP)
Parse Blaise source files (.bla, .inc) into tokens	Y	Y
Keep Comments	Y	N
Keep include file structure	Y	N
Keep whitespace	Y	N

Why would we want a Source Editor System?

- HRS is longitudinal study.
- It's a 'big' application.
- Most of the large scale (bulk) changes have to do with fields. 'Small' changes usually involve a few hundred fields (10% = 581 changes).
- There often are several "small" changes that take place during the CAI preparation for a field period.

What We Learned

- For the 2004 Descriptor update task
  - The few descriptors mentioned turned out to be a 2,700 descriptors change request
  - The merge key that was provided with the descriptor update request information was the DEP field name

Translator Functions

- Convert DEP field names to defined block and field name
- Report duplicate requests for same defined block and field name

Translator

- Need to ‘translate’ DEP fieldname paths to defined block name.

Block Name (Def)	#	ind	DEP Path
BB	41	1	
BB_Born	42	1	SecB.Born
BB_ShowStateList	43	1	SecB.Born.B076_B003_
BB_ShowStateList	43	2	SecB.LivedArea.B078_B047_

Translator

- Several DEP fields update request with one define field

Block (Defined)	Name (Defined)	Existing Descriptor	User Descriptor Request	DEP Field Name
BB_Marry	B066_	MARR YEAR BEG	FIRST MARRIAGE YEAR BEGAN	SecB.Marry[1].B066_
BB_Marry	B066_	MARR YEAR BEG	SECOND MARRIAGE YEAR BEGAN	SecB.Marry[2].B066_
BB_Marry	B066_	MARR YEAR BEG	THIRD MARRIAGE YEAR BEGAN	SecB.Marry[3].B066_
BB_Marry	B066_	MARR YEAR BEG	MARRIAGE YEAR BEGAN -4	SecB.Marry[4].B066_

The System
and
Updating a new language

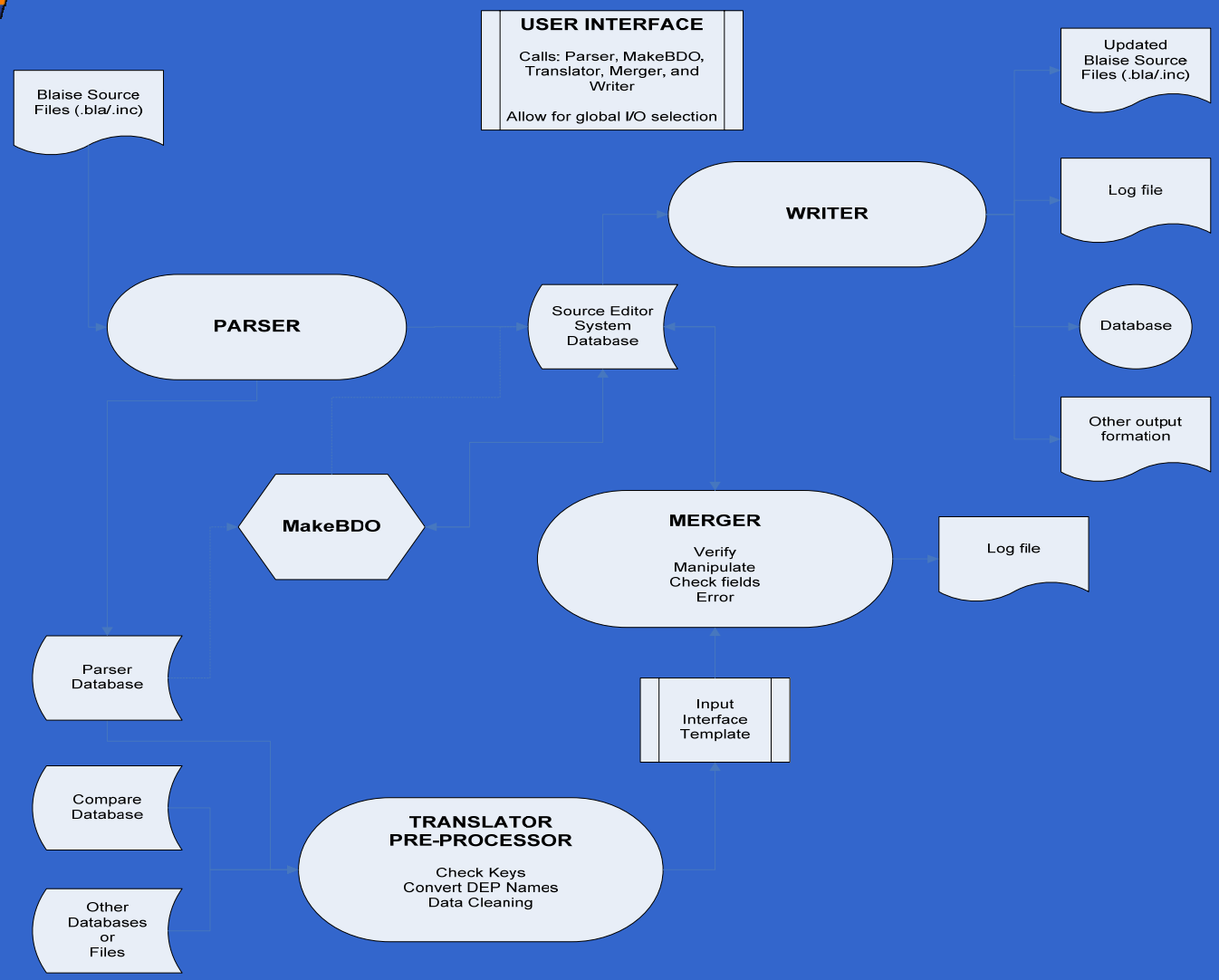
New Language Update

- The early system that handled updating descriptors needed to be expanded to handle the ‘update’ addition of a new language.

System Conceptualization

- Parser Application
- DEP Field Name Translator
- BDO Creation
- Merger Application
- Writer Application
- User Interface

Source Editor System Conceptualization



System Design Considerations

- Encapsulated routines and procedures for each function
- Reusable code versus ad hoc routines
- Tokens described in more meaningful terms
- Language order option
- Parsing whitespace

Language Re-order

Before	Quick	Best
CORE English	CORE English	CORE English
CORE Spanish	CORE Spanish	CORE Spanish
PROXY English	PROXY English	PROXY English
EXIT English	EXIT English	PROXY Spanish
EXIT Spanish	EXIT Spanish	EXIT English
MEDIA	MEDIA	EXIT Spanish
	PROXY Spanish	MEDIA

BDO and Language

- Looking at all statements that can contain language.
  - Field question text
  - enumerated data type code descriptions, and
  - descriptors
- Coming up with a new way to describe all involved tokens, and merge keys for the token types.

Other considerations

- Update request for enumerations (defined at the field)
  - Merge keys needed
  - BDO for an enumeration type statement

Fields

TP50 "What type of fish do you have?": (n "none", f "fresh water", s "salt water")

Other considerations

- Update request for enumerations (defined as a type)
  - Merge keys needed
  - Look and find the type

TP60 "What type of mammals aside from dogs or cats do you have?“ : TMammals

Blkname	Field Name	Type Name	Type No	Blk path	Blk No End	Blk NameEnd	Type NameEnd	TokenType
SE_illustration		tmammals	25	1..	1	SE_illustration	tmammals	TName
B2_Other	TP50		40					
B2_Other	TP60		41	8.1 ..	1	SE_illustration	tmammals	FTName

Other considerations

- **HRS doesn’t use LID – need to put in “” to keep correct relative positioning for languages**

Before

TPosition =

(STANDING (1) "Standing" "Parado",

SITTING (2) "Sitting" "Sentado" ,

LYINGDOWN (3) "Lying down" "Acostado")

After

TPosition =

(STANDING (1) "Standing" "Parado“ {SE Added text} "Standing" {LS} "Parado" ,

SITTING (2) "Sitting" "Sentado“ {SE Added text} "Sitting" {LS} "Sentado" ,

LYINGDOWN (3) "Lying down" "Acostado“ {SE Added text} "Lying down" {LS} "Acostado")

Other considerations

- Report when update request not found in the data
- Report duplicate update request for the same defined field with option to process/not process those records

Current Use & Future Plans

- The time difference between machine versus manual work.
- Is it worth the effort?

Current System Performance

- *8 hours - Parse 175,600 lines of code*
- *2 hours - Type Cross Reference*
- *6 hours - Create BDO Table*
- *6 hours - Merge 6,500 updates*
- *3 minutes - Write files*

Total Time: 22 hours

Other Uses

- Analysis of Datamodel – being able to look at data within context of tokens
- An alternative way of making updates, i.e. one doesn't necessarily need to use the merger. The system allows for direct edits to the tables in the database.

Tasks for the Current System

- Strip out obsolete or dated comments from prior years.
- Update tags. Modify tags to be more descriptive.
- Update descriptors. Modify labels for data out.
- Update data types. Modify field size, field ranges, etc.
- Update language text. Text provided by another system such as a product from the HRS translation group.

System Benefits

- **Time saving, resulting in faster turn-around of tasks.**
- **Hundreds of changes can be made at one time.**
- **More accurate placement of updates and therefore better quality.**
- **May reduce repetitive-use injury.**

System Benefits (2)

- Robust enough to handle applications as large as HRS.
- Generic enough to handle other non-HRS Blaise CAI applications.
- The application can add or re-order languages.
- The application has features to help handle scale issues.

Blaise Source Code Editing System

- Acknowledgments
  - HRS Instrument Development Team

Blaise Source Code Editing System

Questions?

Contact Information:

Danilo Gutierrez
danilog@umich.edu

Sheila Deskins
sld@umich.edu

FAQ

- Q: In the paper you didn't implement a **translator**?
- A: For the language update, ***we gave*** the user a table for fields and a table for enumeration ***with keys needed*** so we wouldn't have to develop the translator, or worry about multiple update request for the same defined field, for this system development round.
- A: We'll look at how we want to develop this implementation in the future.

Question

- Q: What **technology** did you use?
- A:
  - Parser in VB6 (Access or SQLServer)
  - Merger, Writer in VB.NET and ADO.NET
  - Source Editor Database in SQLServer
  - Proof of concept in C# .NET

Miscellaneous:
- Ad hoc routines in SAS
- BCP