Explicitly Representing Superimposed Information in a Conceptual Model

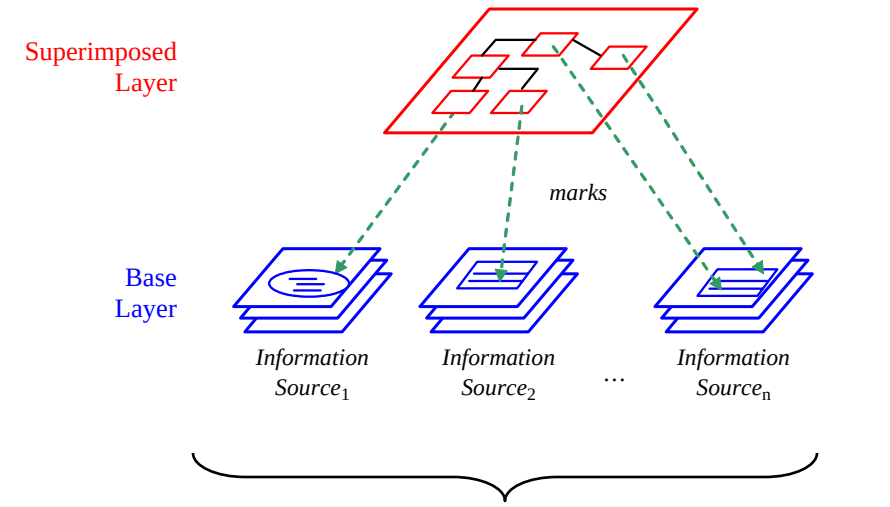
Sudarshan Murthy, Lois Delcambre, David Maier

Department of Computer Science, Portland State University

http://sparce.cs.pdx.edu

mailto:smurthy@cs.pdx.edu

Superimposed Information (SI)



- SI is new information overlaid on existing heterogeneous *base information* (BI)
  - Add new data
  - Impose new schema, model
- *Mark*: A reference to a fragment
- Benefits
  - Create multiple, simultaneous organizations of base information
  - Make new connections among base fragments
  - Preserve context
- A superimposed application (SA) manages SI; base application manages BI

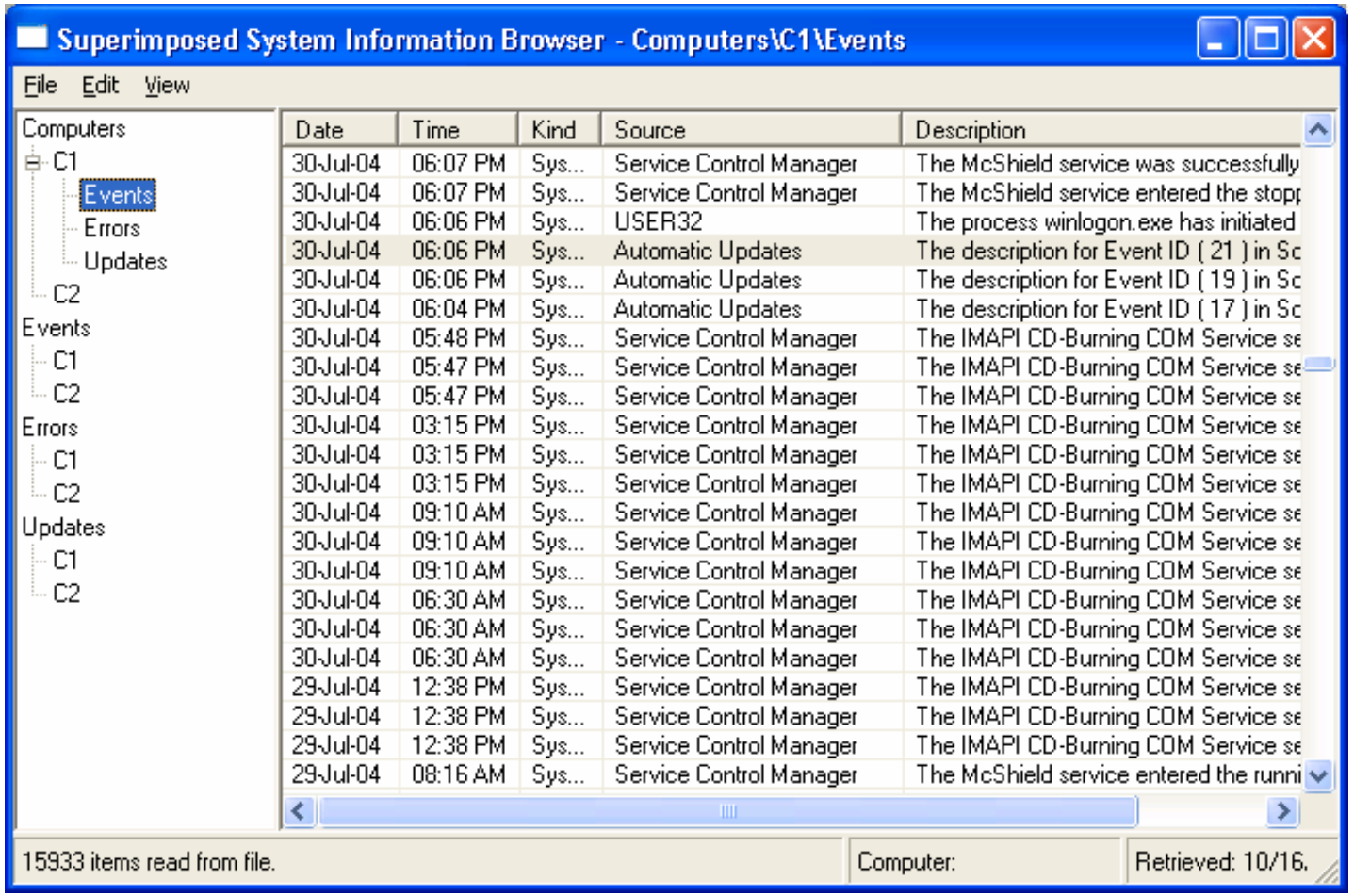
Outline

- SSIB: A superimposed application
- Limitations of the ER model in representing SI
- Extensions to the ER model to represent SI
- Generating relational schemas
- Querying bi-level information
- Summary

The Superimposed System-Information Browser (SSIB)

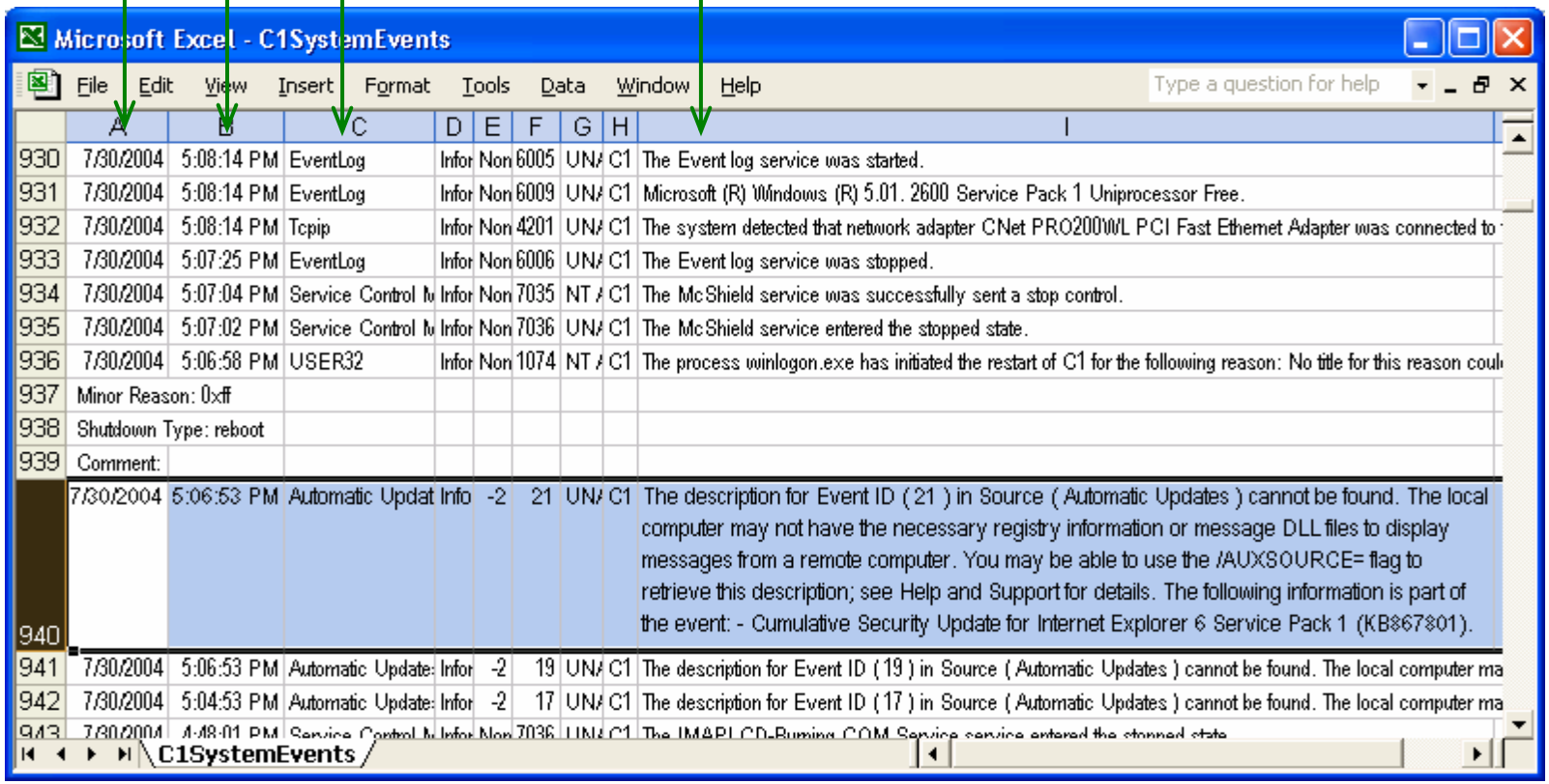
- Allows a system (network) administrator to browse information about computers in a network
  - Application, system, and security events logged
  - Errors recorded/reported
  - Operating-system updates applied
  - Applications installed
  - User observations/comments
- SSIB is one of many SAs we and others have developed

The Browser


Superimposed System Information Browser - Computers\C1\Events
File Edit View
Computers
C1
Events
Errors
Updates
C2
Events
C1
C2
Errors
C1
C2
Updates
C1
C2
Date Time Kind Source Description
30-Jul-04 06:07 PM Sys... Service Control Manager The McShield service was successfully
30-Jul-04 06:07 PM Sys... Service Control Manager The McShield service entered the stopp
30-Jul-04 06:06 PM Sys... USER32 The process winlogon.exe has initiated
30-Jul-04 06:06 PM Sys... Automatic Updates The description for Event ID (21) in Sc
30-Jul-04 06:06 PM Sys... Automatic Updates The description for Event ID (19) in Sc
30-Jul-04 06:04 PM Sys... Automatic Updates The description for Event ID (17) in Sc
30-Jul-04 05:48 PM Sys... Service Control Manager The IMAPI CD-Burning COM Service se
30-Jul-04 05:47 PM Sys... Service Control Manager The IMAPI CD-Burning COM Service se
30-Jul-04 05:47 PM Sys... Service Control Manager The IMAPI CD-Burning COM Service se
30-Jul-04 03:15 PM Sys... Service Control Manager The IMAPI CD-Burning COM Service se
30-Jul-04 03:15 PM Sys... Service Control Manager The IMAPI CD-Burning COM Service se
30-Jul-04 03:15 PM Sys... Service Control Manager The IMAPI CD-Burning COM Service se
30-Jul-04 09:10 AM Sys... Service Control Manager The IMAPI CD-Burning COM Service se
30-Jul-04 09:10 AM Sys... Service Control Manager The IMAPI CD-Burning COM Service se
30-Jul-04 09:10 AM Sys... Service Control Manager The IMAPI CD-Burning COM Service se
30-Jul-04 06:30 AM Sys... Service Control Manager The IMAPI CD-Burning COM Service se
30-Jul-04 06:30 AM Sys... Service Control Manager The IMAPI CD-Burning COM Service se
30-Jul-04 06:30 AM Sys... Service Control Manager The IMAPI CD-Burning COM Service se
29-Jul-04 12:38 PM Sys... Service Control Manager The IMAPI CD-Burning COM Service se
29-Jul-04 12:38 PM Sys... Service Control Manager The IMAPI CD-Burning COM Service se
29-Jul-04 12:38 PM Sys... Service Control Manager The IMAPI CD-Burning COM Service se
29-Jul-04 08:16 AM Sys... Service Control Manager The McShield service entered the runni
15933 items read from file.
Computer:
Retrieved: 10/16.

Event Log

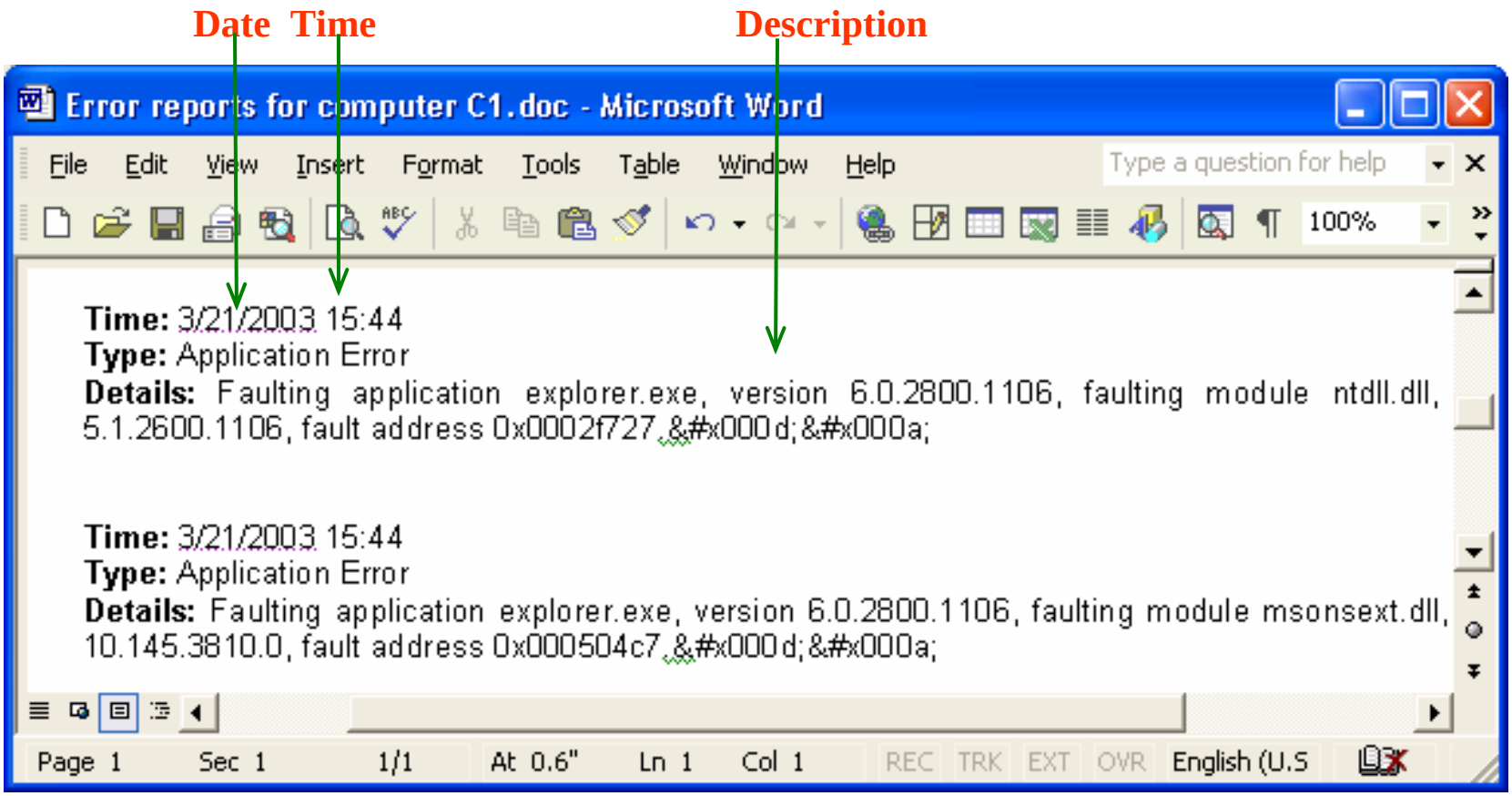
Date **Time** **Source** **Description**



	A	B	C	D	E	F	G	H	I
930	7/30/2004	5:08:14 PM	EventLog	Infor	Non	6005	UNA	C1	The Event log service was started.
931	7/30/2004	5:08:14 PM	EventLog	Infor	Non	6009	UNA	C1	Microsoft (R) Windows (R) 5.01. 2600 Service Pack 1 Uniprocessor Free.
932	7/30/2004	5:08:14 PM	Tcpip	Infor	Non	4201	UNA	C1	The system detected that network adapter CNet PRO200WL PCI Fast Ethernet Adapter was connected to
933	7/30/2004	5:07:25 PM	EventLog	Infor	Non	6006	UNA	C1	The Event log service was stopped.
934	7/30/2004	5:07:04 PM	Service Control M	Infor	Non	7035	NT A	C1	The McShield service was successfully sent a stop control.
935	7/30/2004	5:07:02 PM	Service Control M	Infor	Non	7036	UNA	C1	The McShield service entered the stopped state.
936	7/30/2004	5:06:58 PM	USER32	Infor	Non	1074	NT A	C1	The process winlogon.exe has initiated the restart of C1 for the following reason: No title for this reason coul
937	Minor Reason: 0xff								
938	Shutdown Type: reboot								
939	Comment:								
940	7/30/2004	5:06:53 PM	Automatic Updat	Info	-2	21	UNA	C1	The description for Event ID (21) in Source (Automatic Updates) cannot be found. The local computer may not have the necessary registry information or message DLL files to display messages from a remote computer. You may be able to use the /AUXSOURCE= flag to retrieve this description; see Help and Support for details. The following information is part of the event: - Cumulative Security Update for Internet Explorer 6 Service Pack 1 (KB867801).
941	7/30/2004	5:06:53 PM	Automatic Update:	Infor	-2	19	UNA	C1	The description for Event ID (19) in Source (Automatic Updates) cannot be found. The local computer ma
942	7/30/2004	5:04:53 PM	Automatic Update:	Infor	-2	17	UNA	C1	The description for Event ID (17) in Source (Automatic Updates) cannot be found. The local computer ma
943	7/30/2004	4:48:01 PM	Service Control M	Infor	Non	7036	UNA	C1	The IMAPLCD-Running COM Service service entered the stopped state.

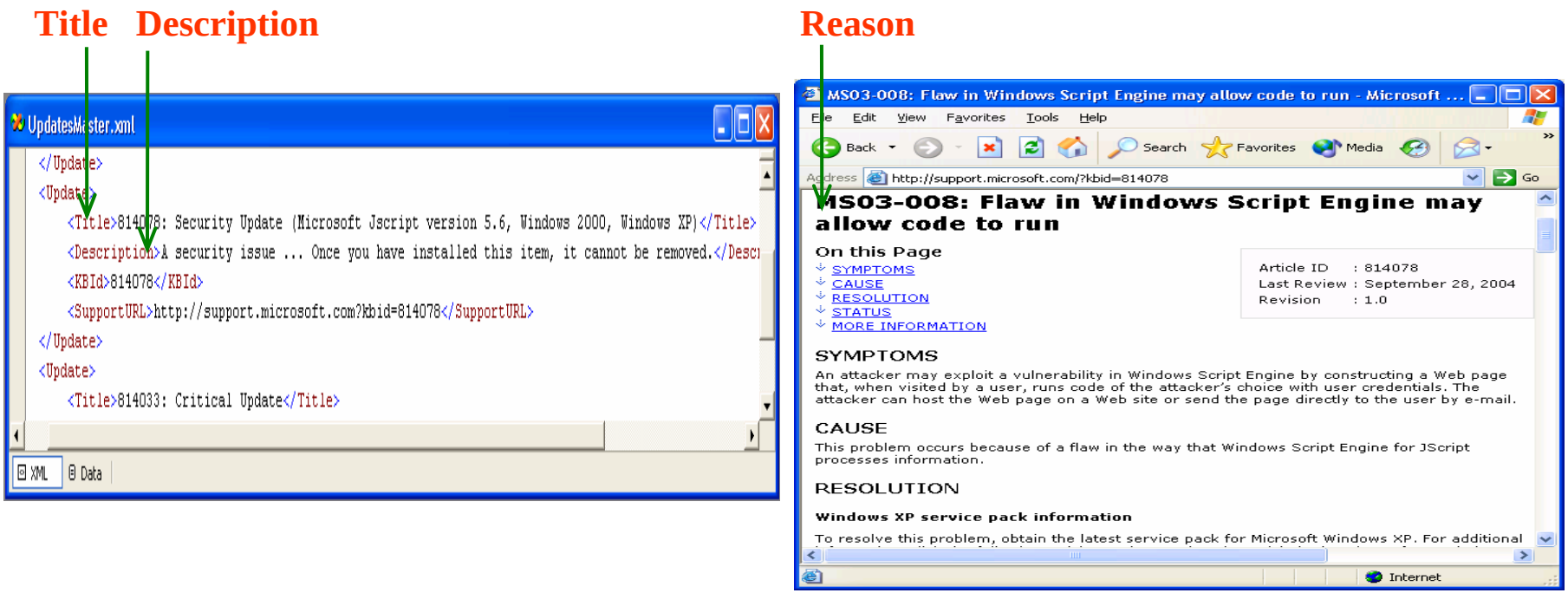
Some structural variations exist, but information is neatly in a table. Date and time are in separate fields

Error Reports



Uniform structure, but mapping is not clean: Date and time are both in the Time field

OS Updates



Data is heterogeneous and distributed: some data in XML, some in HTML

Structure varies: support URL not always defined, HTML page structure varies widely

Benefits from SSIB

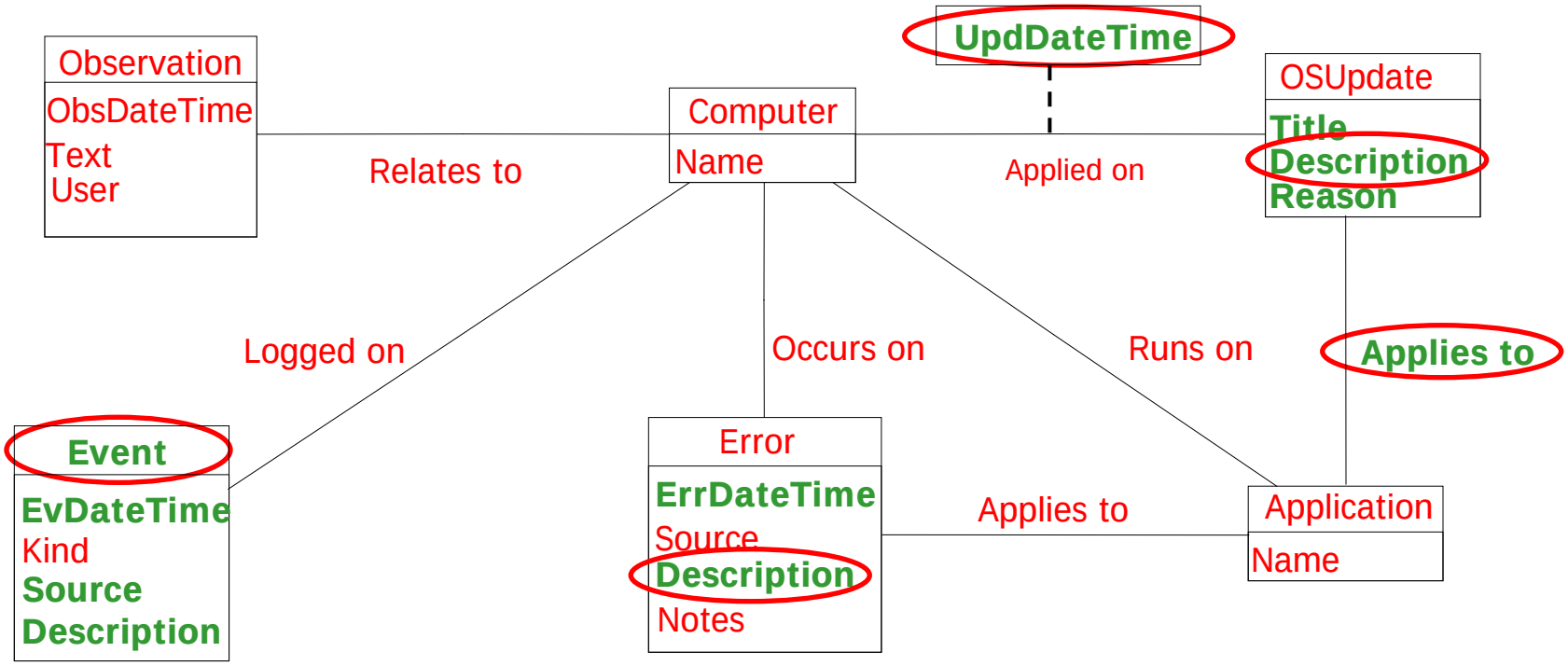
- Puts information in context
  - Relates and elaborates fragments of existing information
- Hides the heterogeneity and distribution of BI
  - Formats: Excel, Word, XML, HTML
  - Locations: Local, network, web
- Allows *structured querying* over structured, semi-structured, and unstructured information
  - Though BI is heterogeneous and distributed, the various data are logically interconnected

Outline

- SSIB: A superimposed application
- Limitations of the ER model in representing SI
- Extensions to the ER model to represent SI
- Generating relational schemas
- Querying bi-level information
- Summary

* All entities have key attribute ID (not shown); all relationships are many-many. Bold (green) text indicates use of marks.

A Conceptual Schema for SSIB*



The Conceptual Modeling Problem

- Traditional modeling cannot represent the use of marks and the presence of BI
- Can model SI and BI separately, but that does not model the cross-over (*i.e.*, marks)
- Can model SI and BI together as if they are one, but that too does not model use of marks
  - We do *not* replicate BI: Marks reference BI *in situ*
- Unable to query across layers without marks
  - Generated logical schema incomplete w.r.t. marks

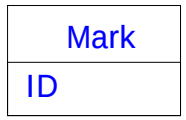
Outline

- SSIB: A superimposed application
- Limitations of the ER model in representing SI
- Extensions to the ER model to represent SI
- Generating relational schemas
- Querying bi-level information
- Summary

Goals and Means

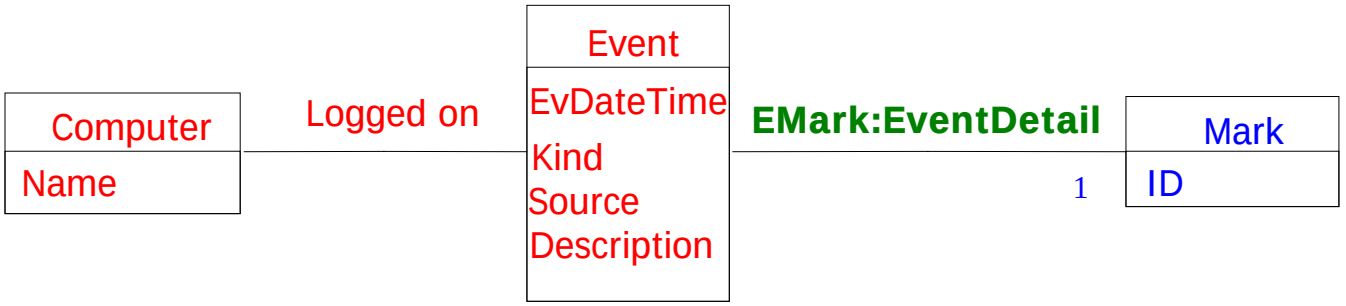
- Goals
  - Associate marks with any schema element
  - Make a value an *excerpt* obtained from base layer
  - Identify and describe contexts of use of marks
  - Allow expression of constraints when using marks
  - Fix the syntax and semantics for/of use of marks
- Means
  - Represent a use of mark as a relationship, and use a convention to represent such relationships
  - Model different uses of marks as *relationship patterns* (recurring problems or needs)

Solution Overview



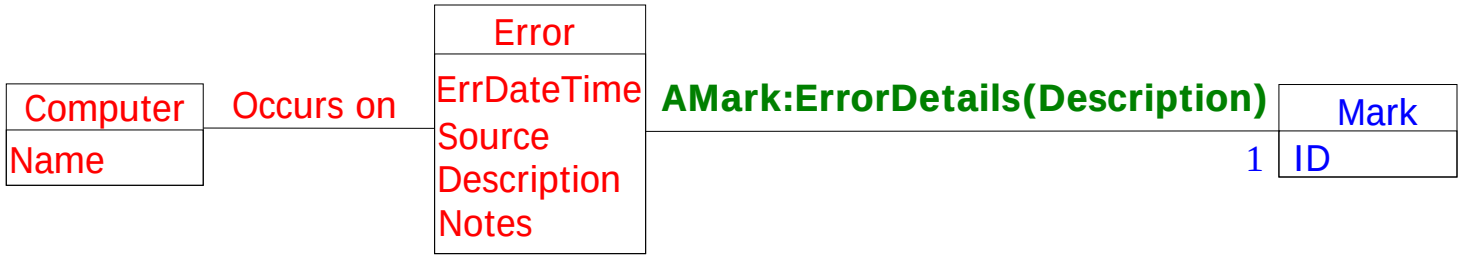
- The `Mark` entity models a mark
  - An `ID` attribute uniquely identifies a mark
  - All marks support the function *resolve*. Some marks also support the function *excerpt*
- The use of a mark is shown as a relationship between this entity and an SI element
  - SI element may be entity, attribute, or relationship
- We identify *five* patterns of using marks

The EMark (Entity-Mark) Pattern



- Signature: EMark:<type>
- Semantics: Associates marks with *entire* entities, not with any particular set of entity-attributes
- Constraints: Any SI entity, any degree, any cardinality constraint

The AMark (Attribute-Mark) Pattern



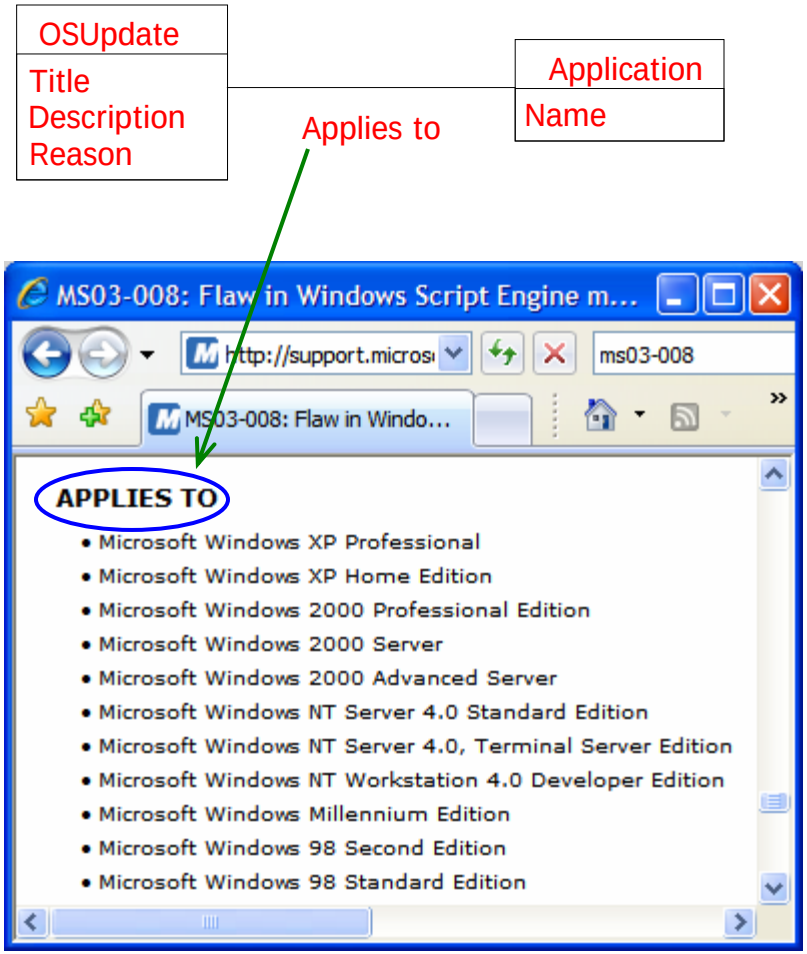
- Signature: AMark:<type>($a_1, a_2, \ldots, a_n$)
- Semantics: All attributes are associated with the same mark; does *not* mean attribute value is obtained from the mark
- Constraints: Any SI attribute, always binary, any cardinality constraint

The AExcerpt (Attribute-Excerpt) Pattern

- Signature: AExcerpt:<type>(a)
- Semantics: Attribute is associated with a mark, *and* its value is the excerpt obtained from the associated mark
- Constraints: Any SI attribute, always binary, cardinality always 1 (assuming single-valued attributes)

* Bowers, *et al*. Superimposed Schematics: Introducing E-R Structure for In-Situ Information Selections. ER 2002.

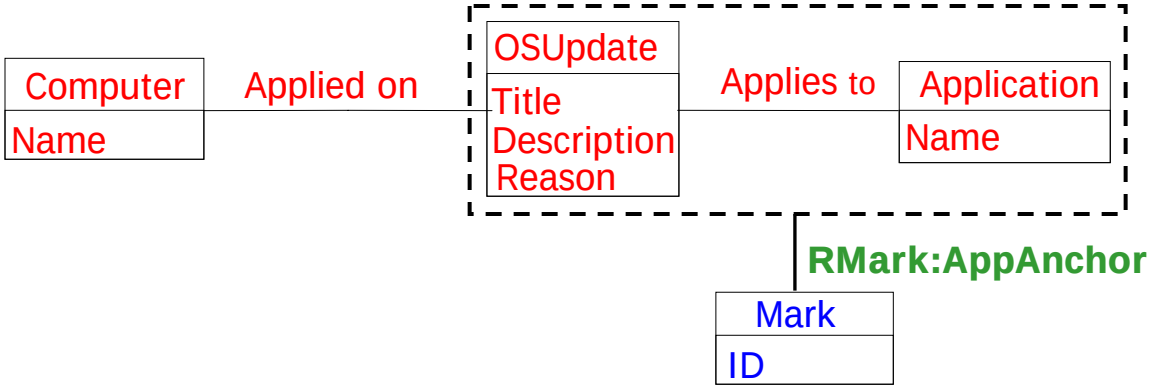
Anchored Relationships



- Some relationships have support or evidence (called *anchors**) in BI
  - *E.g.*, list of applications an update applies to is in a web page

* `Applies to` is the anchored relationship in this example

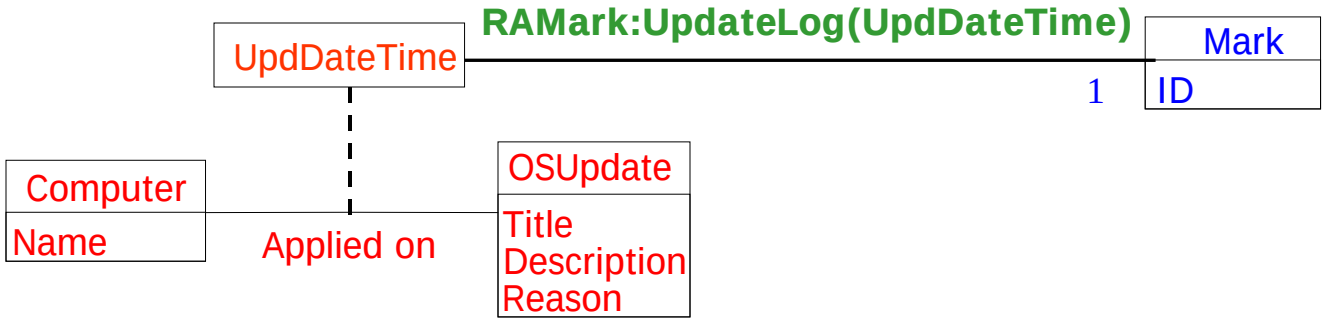
The `RMark` (Relationship-Mark) Pattern



- Signature: `RMark:<type>`
- Semantics: Associates marks with *entire* relationships
- Constraints: No constraints on the anchored relationship*; `RMark` is always binary, any cardinality constraint, may have attributes

* `Applied on` is the anchored relationship in this example

The `RAMark` (Rel. Attribute-Mark) Pattern



- Signature: `RAMark:<type>`($a_1, a_2, \ldots, a_n$)
- Semantics: All attributes are associated with the same marks
- Constraints: No constraints on the anchored relationship*; `RAMark` is always binary, any cardinality constraint, may have attributes

Outline

- SSIB: A superimposed application
- Limitations of the ER model in representing SI
- Extensions to the ER model to represent SI
- Generating relational schemas
- Querying bi-level information
- Summary

Conversion Basics

- The `Mark` entity type is represented as a relation with attributes such as `ID`
  - The exact set of attributes are derived from the definition of the mark abstraction in *SPARCE*, our middleware for SI management
- `EMark` and `AMark` relationship types are converted using the Elmasri and Navathe procedure (*traditional procedure*)
- In the examples we add the key attribute `ID` to relations corresponding to the SI entities

Converting AExcerpt Relationships

- Define a stored relation
  - Convert the relationship type and the related entity types using the traditional procedure
  - *Remove* fields that correspond to attributes whose values will be excerpts
- Define a view
  - Expose attributes not participating in AExcerpt relationship types as they are
  - Use a *user-defined function* (UDF) excerpt to expose attributes involved in AExcerpt relationship types

Example AExcerpt Conversion

* **excerpt** is a UDF

Example AExcerpt Conversion

```
CREATE TABLE Stored_Update
(ID Integer NOT NULL
PRIMARY KEY,
Title VARCHAR(100),
AExcerpt_Desc Integer
NOT NULL REFERENCES
Mark(ID),
~~Description VARCHAR(255),~~
Reason VARCHAR(100),
)
```

```
CREATE VIEW Update
(ID, Title,
 Description,
 Reason) AS
SELECT
ID, Title
excerpt(AExcerpt_Desc),
Reason
FROM Stored_Update
```

Outline

- SSIB: A superimposed application
- Limitations of the ER model in representing SI
- Extensions to the ER model to represent SI
- Generating relational schemas
- Querying bi-level information
- Summary

Bi-level Queries

- A *bi-level query* is a query over SI and BI
- Bi-level queries may be expressed against a conceptual schema that explicitly shows use of marks
- Bi-level queries may be expressed in a language appropriate to the data model
  - SQL for the relational model, XQuery for the XML model, and so on

Example Bi-level Query

- List updates related to security

```
SELECT * FROM Update
WHERE Description LIKE 'Security%'
```

- `Update` is a view, and the value of the attribute `Description` is retrieved from the base layer when the view definition is executed
  - Use only standard SQL to query data both inside and outside a relational database

Outline

- SSIB: A superimposed application
- Limitations of the ER model in representing SI
- Extensions to the ER model to represent SI
- Generating relational schemas
- Querying bi-level information
- Summary

Summary

- Associating marks with entities, attributes, and relationships is a recurring need when modeling SI
- We have identified different patterns of use of marks and extended the ER model
  - We have procedures to create relational schemas from conceptual schemas, and we have shown bi-level queries in standard SQL
- Future work
  - Types of marks, domain-specific BI modeling,
  - Marks in other conceptual modeling frameworks

Q&A

Sudarshan Murthy, Lois Delcambre, David Maier

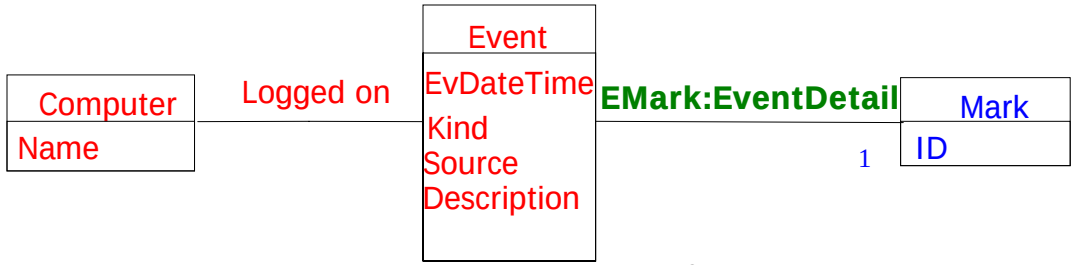
Department of Computer Science, Portland State University

http://sparce.cs.pdx.edu

mailto:smurthy@cs.pdx.edu

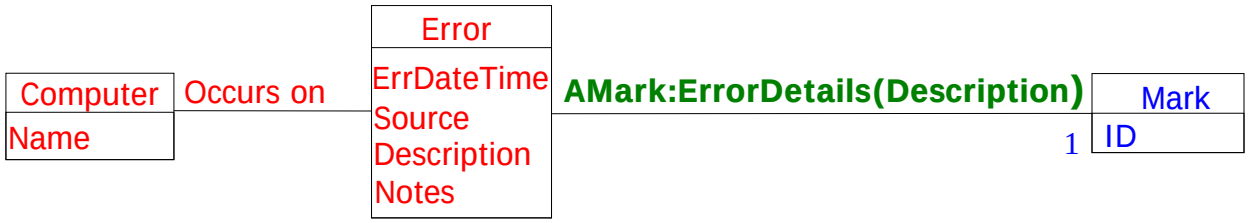
Example EMark Conversion

```
CREATE TABLE Event
( ID Integer NOT NULL PRIMARY KEY,
  EvDateTime Timestamp,
  Kind CHAR(5),
  Source VARCHAR(25),
  Description VARCHAR(255),
  EMark_EventDetail Integer NOT NULL
     REFERENCES Mark(ID)
)
```



Example AMark Conversion

```
CREATE TABLE Error
( ID Integer NOT NULL PRIMARY KEY,
  ErrDateTime Timestamp,
  Source VARCHAR(25),
  Description VARCHAR(255),
  AMark_Error_Desc Integer NOT NULL
     REFERENCES Mark(ID),
  Notes VARCHAR(255)
)
```



* This example allows many marks for the anchored relationship

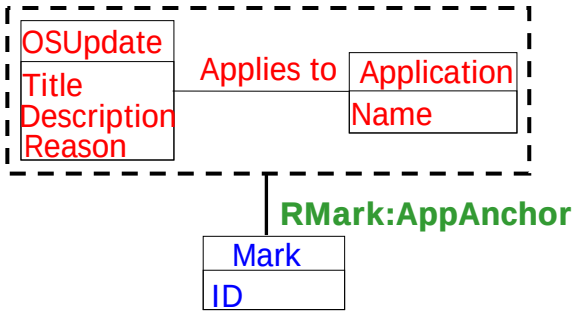
Example RMark Conversion*

```
CREATE TABLE Stored_OSUpdate
(ID Integer NOT NULL PRIMARY KEY, Title…)

CREATE TABLE Application
(ID Integer NOT NULL PRIMARY KEY, Name…)

CREATE TABLE AppliesTo
(UID …, AID …, PRIMARY KEY (UID, AID))

CREATE TABLE RMark_Application
(UID …, AID …,
 RMarkID Integer REFERENCES Mark(ID),
 PRIMARY KEY (UID, AID, RMarkID)
)
```



Example Bi-level Query 2

- Retrieve all errors MS Word caused in the last week

```
SELECT ErrDateTime, Source, Description, Notes
FROM Error
WHERE (ErrDateTime BETWEEN CURRENT_DATE
AND CURRENT_DATE - INTERVAL '6' DAY) AND
(excerpt(AMark_Error_Desc) LIKE '%Word.exe%‘)
```

- The UDF `excerpt` returns error description from the base layer
  - If error description is modeled after the `AExcerpt` pattern, the call to the UDF would not be needed

Example Bi-level Query 3

- A timeline of errors on computer *C2* since the last update related to *MS Outlook* was applied on that computer

```
SELECT ErrDateTime, Description
FROM Error JOIN OccursOn JOIN Computer C
WHERE C.Name = 'C2' AND ErrDateTime > ANY
(SELECT MAX(UpdDateTime)
FROM OSUpdate JOIN AppliesTo JOIN Application A
JOIN AppliedOn
WHERE C.ID = AppliedOn.CID AND A.Name LIKE '%Outlook%')
```

ER Relationships Require Entities

- ER relationships are between entities, but sometimes an attribute carries a mark (*e.g.*, `Error.Description`)
- Promoting attributes to entities, to show relationships, can cause entity proliferation
  - The SSIB schema has *nine* such attributes

Attribute Value

- In ER, no dereferencing is involved in obtaining an attribute's value, but obtaining a value from an attribute that uses a mark may involve dereferencing
  - *E.g.*, `OSUpdate.Description` is the text excerpt for a mark
- Introducing a new domain such as "Mark" does not suffice
  - We need to be able to distinguish between a value that is a mark and a value obtained using a mark