

Running head: TRAVEL MEMORABILIA COLLECTION

Travel Memorabilia Collection: Creating the Database

Zorica Kovacevic, Jessica Teeter, and Kristin Yiotis

San Jose State University

LIBR 202.01

Information Retrieval

Professor: Judy Weedman

March 11, 2004

## PART A

**Alpha Team:** Zorica Kovacevic  
Jessica Teeter  
Kristin Yiotis

**Beta Team:** Tricia Chen  
Yuki Kobayashi  
Sambo Uy

### 1. Statement of Purpose

The purpose of this database is to organize, preserve and make accessible a collection of travel memorabilia of entrance/admission tickets to various popular and cultural places visited by tourists and travelers all over the world.

Users of this collection would be anthropologists, museologists, historians, teachers, writers, journalists, artists, students, tour guides, travelers, or anyone interested in studying print artifacts that represent places with cultural or monetary value and the symbolic and social behaviors associated with those places. [Interesting premise.](#)

By examining this collection, users will be able to connect these material objects to specific places of cultural significance and to understand universal social norms of behavior shared by participants from all over the world.

The database allows the user to search the collection based on the physical characteristics of the objects and information provided on the object such as place name, place type, country, city, language, currency value, content, visual image, logo, date, validation, physical condition.

### 2. Description of the Database

Our data consists of 14 admission tickets necessary for entrance into well-known places of interest all over the world. The tickets themselves become souvenirs that travelers and tourists paste into scrapbooks along with photographs and postcards to remember their journey.

We have examined the tickets looking for attributes that we reworked into fields to create a structure that enabled data entry and search. We chose a single ticket as the unit of analysis since all in our collection were individual, unique units. We needed fields that would aggregate tickets with like qualities and distinguish between unlike ones. Once we created data structure, we created records. Each record has the same structure and the same fields. Only the values that reflect a single entity are specific.

We have decided that certain obvious physical qualities were not useful to our users: such as: size or dimension, the quality of paper, colors evident, the location of the information (face or reverse side). And some otherwise distinguishing features we had to ignore because of our own ignorance, such as translating from foreign languages.

We tried to predict how users would want to search the data and chose fields that permitted grouping data in useful ways as well as separating out unique attributes. One such field, “place name,” designates the name of the place of interest. As each ticket in our collection is unique, this field allows users to segregate one record out of many. All of the other fields, except for the unique ID number, aggregate records according to qualities, such as “place type,” which allows users to collect all tickets from national parks or holy places; “country and city,” all places of interest in Italy or in Florence; or by “content,” within which users could search for all tickets with maps on them, etc. [This is a very good explanation.](#)

In creating the searchable fields and their contents we considered the needs of our users. Tour guides and travelers would want to know what data is available by country. Anthropologists or journalists might want to know what is available by type, such as temple, or ruin. Artists would be interested the design aspects of the tickets, searchable as representative or decorative in the “visual image” field.

We designed the database so that users could draw conclusions about universal norms of social behavior that these tickets indirectly represent. For example: the fact that each ticket is printed with a monetary value makes clear that people must pay for the privilege of visiting places in one country that might be free in another, it indirectly represents the cultural value of these places in the home country. Also, rules of behavior are printed on many tickets. They represent behavioral norms expected of visitors when entering places that are held as sacred or holy by the host people. [Interesting. I wondered as I started reading the paper whether you would be able to create fields that relate directly to culture and social norms – and you were.](#)

### 3. Data Structure and Validation List

Textbase Structure

```
Textbase: C:\Documents and Settings\beba\My Documents\travel
memorabilia\travel_memorabilia
Created: 2/27/2004 5:35:13 PM
Modified: 3/2/2004 6:54:04 PM
```

Field Summary:

1. ID: Automatic Number(next avail=15, increm=1), Term
2. place name: Text, Term & Word  
Validation: required, single-only
3. place type: Text, Term & Word  
Validation: valid-list
4. country: Text, Term & Word  
Validation: single-only
5. city: Text, Term & Word  
Validation: single-only
6. language: Text, Term & Word  
Validation: required
7. currency value: Text, Term & Word

good use of database functionality
---------------------------------------

- Validation: single-only
- 8. content: Text, Term & Word
  - Validation: required, valid-list
- 9. visual image: Text, Term & Word
  - Validation: required, valid-list
- 10. logo: Text, Term & Word
  - Validation: required, single-only, valid-list
- 11. date: Text, Term & Word
  - Validation: single-only
- 12. validation: Text, Term & Word
  - Validation: single-only, valid-list
- 13. physical condition: Text, Term & Word
  - Validation: required, valid-list

Log file enabled, showing 'ID'  
 Leading articles: a an the  
 Stop words: a an and by for from in of the to

Textbase Defaults:  
 Default indexing mode: SHARED IMMEDIATE  
 Default sort order: <none>  
 Textbase passwords:  
 Master password = ''  
 0 Access passwords:  
 No Silent password

Validation list for field 'place type':

church  
 garden  
 museum  
 other  
 palace  
 park  
 ruin  
 temple  
 transportation

Validation list for field 'content':

address  
 cost  
 history  
 hours of operation  
 map  
 none  
 phone number  
 rules  
 serial number

Validation list for field 'visual image':

decorative  
 none  
 representative

Validation list for field 'logo':

no  
yes

Validation list for field 'validation':

no  
yes

Validation list for field 'physical condition':

excellent  
fair  
good  
poor

#### 4. Rules for Indexing

The unit of analysis for this database is one individual ticket. Each ticket is assigned an identification number.

Mandatory fields: ID, place name, language, content, visual image, logo, physical condition, logo

Repeatable fields: Place type, language, content, visual image, date

Use American spelling.

1. ID – (auto-numeric)
2. Place name – (free text)
  - 2.1.RULE: spell out as it appears on the ticket
  - 2.2.RULE: Definition of place: the given name of the attraction, exhibit, point of interest, etc.
3. Place type – museum, temple, garden, ruin, transportation, park, other
  - 3.1.RULE: Definition of place: the attraction, exhibit, point of interest, etc.
  - 3.2.RULE: Definition of type: the category of the attraction, exhibit, point of interest, etc.
4. Country – (free text)
  - 4.1.RULE: Spell out full name, using American spelling
  - 4.2.RULE: If given on ticket, enter; if verified from an external source, enter; if unable to verify, leave blank. Acceptable external sources include: maps, encyclopedias, etc.
5. City – (free text)
  - 5.1.RULE: Spell out full name, use American spelling
  - 5.2.RULE: If given on ticket, enter; if verified from an external source, enter; if unable to verify, leave blank. Acceptable external sources include: maps, encyclopedias, etc. **Province ?**
6. Language – (free text)
  - 6.1.RULE: spell out full name with American spelling
7. Currency value – (free text)
  - 7.1.RULE: original currency amount in numerals and the currency type spelled out in American English **Does “type” mean the unit – dollar, peso, pound, etc.? If so, what is the American English spelling of “peso”? I think this may need some rewording.**
  - 7.2.RULE: enter "free" if no fee is charged
8. Content – map, hours of operation, serial #, history, address, phone number, rules, none
  - 8.1.RULE: Definition of history: historical information about the attraction, exhibit, point of interest, etc. **It's useful to indicate in your rules (even though it's built into the data structure) whether or not more than one value can be assigned.**

9. Visual image – representative, decorative, none
  - 9.1.RULE: Definition: representative – shows the actual attraction **or a part of it?**; decorative – decorates the object
10. Logo – yes/no
11. Date – (free text)
  - 11.1.RULE: 4 digit year/2 digit month/2 digit day (separated by slashes)
  - 11.2.RULE: Include time of day if listed, with am/pm if indicated **Periods in a.m. or no periods? Space between numbers and am/pm or not?**
12. Validation – yes/no
  - 12.1.RULE: Definition of validation: identifying stamp, punch, mark or tear to indicate whether the ticket is new or used
13. Physical condition – poor, fair, good, excellent
  - 13.1.RULE: Definitions: poor – ripped, dirty; fair – folded, wrinkled, or bent; good – not ripped, folded, bent, or dirty; excellent – new, unused **So only an unused ticket can be in excellent condition, correct?**
  - 13.3.RULE: Definition of ripped vs. torn: Ripped is unintentional damage or wear and tear; torn is a perforated section that has been removed to indicate that the ticket is no longer valid.

## 5. Alpha Team Records

### ID 1

**place name** cable car  
**place type** other  
**country** United States  
**city** San Francisco  
**language** English  
**currency value** 2 dollars  
**content** serial number  
history  
**visual image** representative  
**logo** yes  
**validation** no  
**physical condition** excellent

### ID 2

**place name** Splendid China miniature scenic spot  
**place type** other  
**country** China  
**city** Shenzhen  
**language** Chinese  
English  
**currency value** 60 hong kong dollars  
**content** serial number  
**visual image** representative  
**logo** yes  
**validation** yes  
**physical condition** good

### ID 3

**place name** Trem do Corcovado  
**place type** transportation  
**country** Brasil  
**city** Rio de Janeiro  
**language** Portuguese  
**content** address  
serial number  
**visual image** representative  
**logo** no  
**date** 2003/05/26 11:30  
**validation** no  
**physical condition** good

**ID 4****place name** Zona Arqueologica de Tulum**place type** ruin**country** Mexico**city** Tulum**language** Spanish**currency value** 35.00 Mexican dollars**content** serial number**visual image** representative  
decorative**logo** no**date** 2002/12/31**validation** yes**physical condition** good**ID 5****place name** Basilica Cattedrale di Parma**place type** church**country** Italy**city** Parma**language** Italian**currency value** 5,000 lire**content** address**visual image** representative**logo** yes**validation** no**physical condition** fair**ID 6****place name** Rijks Museum**place type** museum**country** Holland**city** Amsterdam**language** Dutch

English

**currency value** 25.00 gulden**content** address

hours of operation

serial number

**visual image** decorative**logo** no**date** 2000/05/30 10:43**validation** yes**physical condition** good

**ID 7****place name** Prazska Loreta**place type** church**country** Czech Republic**city** Prague**language** English

Czech

German

French

Italian

Spanish

**content** history**visual image** decorative  
representative**logo** no**validation** no**physical condition** fair**ID 8****place name** Museum of the Western Han Dynasty**place type** museum**country** China**city** Guangzhou**language** Chinese

English

**currency value** 5.00 yuen**content** serial number

address

map

phone number

**visual image** representative  
decorative**logo** yes**validation** yes**physical condition** good**ID 9****place name** Santa Sofia Museum**place type** museum**country** Turkey**city** Istambul**language** Turkish

English

**currency value** free**content** rules**visual image** representative**logo** yes**date** 2003/11/14 11:45:15**validation** yes**physical condition** fair

**ID** 10

**place name** Tanah Lot

**place type** temple

**country** Indonesia

**city** Tabanan-Bali

**language** English  
Bahasa

**currency value** 1000 rupiah

**content** serial number

**visual image** decorative

**logo** yes

**date** 1995/02/27

**validation** yes

**physical condition** fair

This – and lire, gulden, etc. – are what I had in mind when I asked about “American English spelling” for the currency unit.

**ID** 11

**place name** Sokkul-am Grotto

**place type** temple

**country** South Korea

**language** Korean  
English

**currency value** 1,500 won

**content** serial number  
history

**visual image** representative

**logo** yes

**date** ----/10/09

**validation** yes

**physical condition** good

**ID** 12

**place name** Galleria degli Uffizi

**place type** museum

**country** Italy

**city** Florence

**language** Italian  
English

**currency value** 6.20 euro

**content** serial number  
rules

**visual image** decorative

**logo** no

**date** 1999/02/13

**validation** yes

**physical condition** good

**ID** 13

**place name** United States Grand Prix

**place type** other

**country** United States

**city** Indianapolis

**language** English

**currency value** 85.00 U. S. dollars

**content** map  
rules

**visual image** decorative

**logo** yes

**date** 2001/09/30

**validation** yes

**physical condition** good

**ID** 14

**place name** Heian Shrine

**place type** temple

**country** Japan

**city** Kyoto

**language** Japanese  
English

**content** hours of operation

**visual image** representative

**logo** no

**validation** yes

**physical condition** fair

## 6. Beta Team Records

### ID 1

**place name** San Francisco Municipal Railway

**place type** transportation

**country** U.S.A

**city** San Francisco

**language** English

**currency value** 2 Dollars

**content** Serial #: M418110. Rules: The ticket is not valid for ride if receipt stub is detached. Valid for a one way cable car trip. History: Invented in San Francisco, the cable cars have been operating here since 1878. Once a form of transportation used in cities around the world, the San Francisco cable cars are the last in daily revenue service. They were made a National Historic Landmark in 1964.

**visual image** decorative

**logo** yes

**validation** no

**physical condition** excellent

### ID 2

**place name** Splendid China Miniature Scenic Spot

**place type** park

**country** China

**city** Shenzhen

**language** Chinese/English

**currency value** 60 HK Dollars

**content** Serial #: 843083.

**visual image** decorative

**logo** no

**validation** yes

**physical condition** good

### ID 3

**place name** Trem do Corcovado

**place type** transportation

**country** Brazil

**city** Rio de Janeiro

**language** Portuguese/English

**content** Serial #: 03535405. Rules: Keep to return. Phone: (0055) (21) 2558-1239.

**visual image** decorative

**logo** yes

**date** 2003/05/26

**validation** yes

**physical condition** good

**ID 4****place name** Zona de Arqueologica de Tulum**place type** ruin**country** Mexico**city** Tulum**language** Spanish**currency value** 35 Dollars**content** Serial #: 1088647.**visual image** decorative**logo** no**date** 2002/12/31**validation** yes**physical condition** good**ID 5****place name** Basilica Cattedrale di Parma Fabbriceria della Basilica Cattedrale**place type** temple**country** Italy**city** Parma**language** Italian**currency value** 5.000 Lire**content** none**visual image** representative**logo** no**validation** no**physical condition** good**ID 6****place name** RIJKS MUSEUM Amsterdam**place type** museum**country** The Netherlands**city** Amsterdam**language** Dutch/English**currency value** 25,00 Guilders**content** Serial #: 200010 001002. Hours: Open daily from 10 A.M till 5 P.M.  
except on New Year's Day. Address: Rijksmuseum Amsterdam,  
Stadhouderskade 42, Amsterdam**visual image** decorative**logo** no**date** 2000/05/30 10:43 04**validation** yes**physical condition** good

**ID 7****place name** Prazska Loreta**place type** temple**country** Czech Republic**city** Prague**language** Czech/English/German/Italian/French/Spanish**content** Since the 17th Century people have come to pray in this place of pilgrimage and have admired the baroque beauty of this structures as well as the treasures enshrined there.**visual image** representative**logo** no**validation** no**physical condition** fair**ID 8****place name** Museum of the Western Han Dynasty, Museum of the Nanyue King  
Guangzhou**place type** temple**country** China**city** Guangzhou**language** Chinese/English**currency value** 5.00 Yuan Renminbi**content** Serial #: 130638. Floor plan of the museum. Address in Chinese.  
Phone: 664920.**visual image** decorative**logo** yes**validation** yes**physical condition** good**ID 9****place name** Santa Sofia Museum**place type** museum**country** Turkey**city** Constantinople-Istanbul**language** Turkish/English**content** Serial #: 2 --- 017. Rules: Valid only in Santa Sofia Museum, professional camera recording is subject to a fee, tickets are non-refundable, keep ticket during the visit, ticket is valid only for one person and only for one time for the day sold.**visual image** representative**logo** no**date** 2003/11/14 11:45:15**validation** yes**physical condition** good

**ID** 10

**place name** Taman Wisata "Tanah Lot"

**place type** temple

**country** Indonesia

**city** Tabanan -- Bali

**language** Indonesian/English

**currency value** 1000 Rupiahs

**content** Serial #: 13856

**visual image** none

**logo** yes

**validation** yes

**physical condition** fair

**ID** 11

**place name** Sokkul-Am Grotto

**place type** temple

**country** Korea

**city** Gyeongju

**language** Korean/English

**currency value** 1,500 won

**content** Serial #: 0168518. This cave temple was built in 751. The numerous stone Buddhist figures in the grotto represent the best sculpture of the Unified Shilla Kingdom (668-935) and of all Korea.

**visual image** representative

**logo** no

**validation** yes

**physical condition** good

**ID** 12

**place name** Galleria degli Uffizi

**place type** museum

**country** Italy

**city** Florence

**language** Italian/English

**currency value** 6.20 Euro

**content** Serial #: 00041337. Entry time: 9:45-10:00. Rules: The ticket can be used only on the issuing date; in case of booking ticket, it has to be used on the booked date.

**visual image** decorative

**logo** no

**date** 1999/02/13

**validation** yes

**physical condition** good

**ID** 13

**place name** Formula 1

**place type** other

**country** U.S.A

**city** Indianapolis

**language** English

**currency value** 85 Dollars

**content** Map of the racetrack. Standing start 1:00 P.M. EST. Rules: Patrons must have possession of ticket at all times.

**visual image** decorative

**logo** yes

**date** 2001/09/30

**validation** yes

**physical condition** good

**ID** 14

**place name** Heian Shrine

**place type** temple

**country** Japan

**city** Kyoto

**language** Japanese

**content** Hours: 8:30 A.M -- 4:30 P.M. Rules: No smoking in the garden (smoking is allowed in the designated area). Heian Shrine is a representative of Meiji Shrine. Garyu bridge was made in 1582. Taiheikaku was built in 1912.

**visual image** decorative

**logo** no

**validation** yes

**physical condition** good

## 7. Comparison (differences are in red)

### Alpha Team Records

**ID** 1  
**place name** cable car  
**place type** other  
**country** United States  
**city** San Francisco  
**language** English  
**currency value** 2 dollars  
**content** serial number  
                   history  
**visual image** representative  
**logo** yes  
**validation** no  
**physical condition** excellent

**ID** 2  
**place name** Splendid China miniature scenic spot  
**place type** other  
**country** China  
**city** Shenzhen  
**language** Chinese  
                   English  
**currency value** 60 hong kong dollars  
**content** serial number  
**visual image** representative  
**logo** yes  
**validation** yes  
**physical condition** good

**ID** 3  
**place name** Trem do Corcovado  
**place type** transportation  
**country** Brasil  
**city** Rio de Janeiro  
**language** Portuguese  
**content** address  
                   serial number  
**visual image** representative  
**logo** no  
**date** 2003/05/26 11:30  
**validation** no  
**physical condition** good

### Beta Team Records

**ID** 1  
**place name** San Francisco Municipal Railway  
**place type** transportation  
**country** U.S.A  
**city** San Francisco  
**language** English  
**currency value** 2 Dollars  
**content** Serial #: M418110. Rules: The ticket is  
                   not valid for ride if receipt stub is  
                   detached. Valid for a one way cable car  
                   trip. History: Invented in San Francisco,  
                   the cable cars have been operating here  
                   since 1878. Once a form of transportation  
                   used in cities around the world, the San  
                   Francisco cable cars are the last in daily  
                   revenue service. They were made a  
                   National Historic Landmark in 1964.  
**visual image** decorative  
**logo** yes  
**validation** no  
**physical condition** excellent

**ID** 2  
**place name** Splendid China Miniature Scenic Spot  
**place type** park  
**country** China  
**city** Shenzhen  
**language** Chinese/English  
**currency value** 60 HK Dollars  
**content** Serial #: 843083.  
**visual image** decorative  
**logo** no  
**validation** yes  
**physical condition** good

**ID** 3  
**place name** Trem do Corcovado  
**place type** transportation  
**country** Brazil  
**city** Rio de Janeiro  
**language** Portuguese/English  
**content** Serial #: 03535405. Rules: Keep to return.  
                   Phone: (0055) (21) 2558-1239.  
**visual image** decorative  
**logo** yes  
**date** 2003/05/26  
**validation** yes  
**physical condition** good



ID 7

**place name** Prazska Loreta**place type** church**country** Czech Republic**city** Prague**language** English

Czech

German

French

Italian

Spanish

**content** history**visual image** decorative  
representative**logo** no**validation** no**physical condition** fair

ID 8

**place name** Museum of the Western Han Dynasty**place type** museum**country** China**city** Guangzhou**language** Chinese

English

**currency value** 5.00 yuen**content** serial number

address

map

phone number

**visual image** representative  
decorative**logo** yes**validation** yes**physical condition** good

ID 9

**place name** Santa Sofia Museum**place type** museum**country** Turkey**city** Istambul**language** Turkish

English

**currency value** free**content** rules**visual image** representative**logo** yes**date** 2003/11/14 11:45:15**validation** yes**physical condition** fair

ID 7

**place name** Prazska Loreta**place type** temple**country** Czech Republic**city** Prague**language** Czech/English/German/Italian/French/Spanish**content** Since the 17th Century people have come to pray in this place of pilgrimage and have admired the baroque beauty of this structures as well as the treasures enshrined there.**visual image** representative**logo** no**validation** no**physical condition** fair

ID 8

**place name** Museum of the Western Han Dynasty,  
Museum of the Nanyue King Guangzhou**place type** temple**country** China**city** Guangzhou**language** Chinese/English**currency value** 5.00 Yuan Renminbi**content** Serial #: 130638. Floor plan of the museum.  
Address in Chinese. Phone: 664920.**visual image** decorative**logo** yes**validation** yes**physical condition** good

ID 9

**place name** Santa Sofia Museum**place type** museum**country** Turkey**city** Constantinople-Istanbul**language** Turkish/English**content** Serial #: 2 --- 017. Rules: Valid only in Santa Sofia Museum, professional camera recording is subject to a fee, tickets are non-refundable, keep ticket during the visit, ticket is valid only for one person and only for one time for the day sold.**visual image** representative**logo** no**date** 2003/11/14 11:45:15**validation** yes**physical condition** good

**ID** 10  
**place name** Tanah Lot  
**place type** temple  
**country** Indonesia  
**city** Tabanan-Bali  
**language** English  
 Bahasa  
**currency value** 1000 rupiah  
**content** serial number  
**visual image** decorative  
**logo** yes  
**date** 1995/02/27  
**validation** yes  
**physical condition** fair

**ID** 11  
**place name** Sokkul-am Grotto  
**place type** temple  
**country** South Korea  
**language** Korean  
 English  
**currency value** 1,500 won  
**content** serial number  
 history  
**visual image** representative  
**logo** yes  
**date** ----/10/09  
**validation** yes  
**physical condition** good

**ID** 12  
**place name** Galleria degli Uffizi  
**place type** museum  
**country** Italy  
**city** Florence  
**language** Italian  
 English  
**currency value** 6.20 euro  
**content** serial number  
 rules  
**visual image** decorative  
**logo** no  
**date** 1999/02/13  
**validation** yes  
**physical condition** good

**ID** 10  
**place name** Taman Wisata "Tanah Lot"  
**place type** temple  
**country** Indonesia  
**city** Tabanan -- Bali  
**language** Indonesian/English  
**currency value** 1000 Rupiahs  
**content** Serial #: 13856  
**visual image** none  
**logo** yes  
**validation** yes  
**physical condition** fair

**ID** 11  
**place name** Sokkul-Am Grotto  
**place type** temple  
**country** Korea  
**city** Gyeongju  
**language** Korean/English  
**currency value** 1,500 won  
**content** Serial #: 0168518. This cave temple was built in 751. The numerous stone Buddhist figures in the grotto represent the best sculpture of the Unified Shilla Kingdom (668-935) and of all Korea.  
**visual image** representative  
**logo** no  
**validation** yes  
**physical condition** good

**ID** 12  
**place name** Galleria degli Uffizi  
**place type** museum  
**country** Italy  
**city** Florence  
**language** Italian/English  
**currency value** 6.20 Euro  
**content** Serial #: 00041337. Entry time: 9:45-10:00. Rules: The ticket can be used only on the issuing date; in case of booking ticket, it has to be used on the booked date.  
**visual image** decorative  
**logo** no  
**date** 1999/02/13  
**validation** yes  
**physical condition** good

**ID** 13  
**place name** United States Grand Prix  
**place type** other  
**country** United States  
**city** Indianapolis  
**language** English  
**currency value** 85.00 U. S. dollars  
**content** map  
           rules  
**visual image** decorative  
**logo** yes  
**date** 2001/09/30  
**validation** yes  
**physical condition** good

**ID** 14  
**place name** Heian Shrine  
**place type** temple  
**country** Japan  
**city** Kyoto  
**language** Japanese  
           English  
**content** hours of operation  
**visual image** representative  
**logo** no  
**validation** yes  
**physical condition** fair

**ID** 13  
**place name** Formula 1  
**place type** other  
**country** U.S.A  
**city** Indianapolis  
**language** English  
**currency value** 85 Dollars  
**content** Map of the racetrack. Standing start 1:00 P.M.  
           EST. Rules: Patrons must have possession of  
           ticket at all times.  
**visual image** decorative  
**logo** yes  
**date** 2001/09/30  
**validation** yes  
**physical condition** good

**ID** 14  
**place name** Heian Shrine  
**place type** temple  
**country** Japan  
**city** Kyoto  
**language** Japanese  
**content** Hours: 8:30 A.M -- 4:30 P.M. Rules: No  
           smoking in the garden (smoking is allowed in the  
           designated area). Heian Shrine is a representative  
           of Meiji Shrine. Garyu bridge was made in 1582.  
           Taiheikaku was built in 1912.  
**visual image** decorative  
**logo** no  
**validation** yes  
**physical condition** good

Interesting differences between the two sets of records.

Part A is very well done! It's clearly written and professionally presented.

## Part B

### Alpha team:

Zorica Kovacevic  
 Jessica Teeter  
 Kristin Yiotis

### Evaluation of Statement of Purpose

Unless we can prove how our database allows users to study the symbolic and social behaviors associated with places of interest, we shouldn't claim it does in our statement of purpose. Also unless we can show how these admission tickets enable users to understand cultural and social norms shared by visitors and host people at these places of interest, we shouldn't claim they can.

Universal norms of behavior that these tickets indirectly represent:

1. Visitors pay to enter places of great cultural/historical/social/popular value. Concepts of sacred space and profane space are universal. Sacred spaces cost to enter, while profane spaces do not. *Are you sure this is true? Disneyland costs to enter, but many churches do not. Unless you're defining sacred space as space which one must pay to enter, this would seem to be a counter-example.*
2. Concepts of public and private space are relative; what is free in one country might cost in another.
3. When people pay to enter they agree to rules. Rules are sometimes written on the ticket, such as what can and cannot be photographed, no smoking, no food, etc. Rules, such as customs, are unwritten, e.g., remove shoes in a temple or a mosque, cover your head in churches, your arms and legs in mosques, etc.
4. Values of a culture can be determined by the amounts of money charged at cultural/historical/social/popular places. And comparisons of prices across cultures can tell something about allocation of resources. For example: in the USA the same person will pay \$85.00 for the Grand Prix, \$2.00 for the SF cable car, \$10.00 for the Hoover Dam, and in Australia \$65.00 at the Sydney Opera House. In China, foreigners aren't supposed to possess local currency, called renminbi, or "people's currency." They pay either in HK dollars or in a special

“foreigner’s currency” and are charged double the fixed government price.  
[Interesting.](#)

### **Evaluation of the Data Structure**

We chose the single ticket as the unit of analysis since all in our collection were individual, unique units. And we think the fields successfully characterized this collection. However, to accommodate an expanded collection, we would have to make additional decisions and possibly rethink the structure. [True.](#) For example: would we enter, as a single record, books of tickets to places of interest with multiple attractions, or would our unit of analysis eliminate these from the collection?

Also, we would need to define strong and weak interpretations of the class admission tickets to determine in what directions the collection could expand. Should we define ahead of time the exact parameters of the collection or take a less strict view of what can be included? Can the class admission tickets expand to include one-time events, such as the opera, theater, the Grand Prix car race, the NFL Rose Bowl, a single ferry ride, tram ride? Or should the class be limited to places where the “event” is continuous, such as a museum, park, or observation tower? We have held to the less strict view in choosing our own data, but have done so arbitrarily and not uniformly. For example: we agreed to include the Grand Prix car race, but to exclude the Sydney Opera because it was out of class. A stricter observance of the class would have eliminated both. [An interesting and important issue.](#)

### **Evaluation of the Rules**

Our database and documentation were not tested sufficiently before the release for beta testing. We did not realize how insufficient our documentation really was. In the interest of time, the same member of the team who entered the fields into DBTextWorks also keyed in the values for the alpha test. If another group member less immediately familiar with the data structure had done the alpha test, our alpha test might have revealed the major problems with the rules. In addition, after the alpha test we should have checked our records against our rules to confirm that the rules were producing the results we expected. If not, we could have refined the rules before releasing the database to the beta testers. The fact that we did not always follow our own rules when entering data resulted in errors in the records that could have been avoided. [All of these are good points.](#) This written evaluation is the first examination of results that we have done.

Revisions are in red; comments and evaluation are in blue.

### **Revised Fields, Descriptions, and Rules** *Travel Memorabilia – Admission Tickets*

The unit of analysis for this database is one individual ticket.

Each ticket is assigned an identification number.

Mandatory fields: ID, place name, language, content, visual image, physical condition

Repeatable fields: Place type, language, content, visual image, date

Use American English spelling **and grammar unless stated otherwise in rules.** Capitalize value **entrees** entries only when proper nouns, such as country, city, place names, street names. **Don't capitalize currencies, dollars, lira, won.**

Does uppercase, lowercase make any difference to DBTextWorks in the search potential for values? If not why worry about them? DBT is not case sensitive for searching. Why worry about them? -- consistency, professional appearance.

We found we should have included examples of rules to clarify them. Yes, this can be extremely helpful.

1. ID – (auto-numeric)
  - 1.1 **RULE: Start with number 1.**

This seems self-evident but DBTextWorks labels the first record as the “next,” which an indexer might think means record #2.

2. Place Name – (free text)
  - 2.1.RULE: Spell out as it appears on the ticket, **such as not capitalized or all caps.**
  - 2.2.RULE: Definition of place: the given name of the attraction, exhibit, point of interest, etc.
  - 2.3 **RULE: Definition of name: 1<sup>st</sup> choose the name in the largest font; if that does not determine name, choose the name in the bold font, 3<sup>rd</sup> choice is the name in quotes or italics, 4<sup>th</sup> choice is the name that is topmost. This is a good “decision tree” – but instead of “choose,” a better verb might be “consider.” “Choose” implies actually selecting that term.**
  - 2.4 **RULE: If you put more than one name, separate by a slash (/), so each name is searchable. F7 makes a better separation to keep both searchable (creates multiple values).**

We did not give an initial assistance in helping the indexer decide the place’s name, most of which are in foreign lands. We needed a rule that established a priority that defined name-- which is printed in the largest font, the bold font, in italics or in quotes, or the topmost name, in that order.

2.4 is a fall back rule in case 2.3 does not fulfill all possibilities. Good idea.

3. Place Type--**Validation List:** museum, temple, garden, ruin, transportation, park, **religious place (includes church, shrine, mosque, burial ground), amusement park, sports venue, other.**
  - 3.1.RULE: Definition of place: the attraction, exhibit, point of interest, etc.
  - 3.2.RULE: Definition of type: the category of the attraction, exhibit, point of interest, etc.

We needed to have more granularity to the values in this validation list to allow for a finer analysis of type, or move into a less granular and more general analysis. For example: the less granular category of Religious Place or Holy Place could include the more granular temple, church, shrine, mosque, etc., subcategories.

We did not always follow our own rules. In some instance we never made necessary changes, e.g., #1, cable car, was entered as “other.” Later we added the value “transportation” and forgot to change “other” to “transportation.” This field should be #2 in sequence coming before, not after, Place Name. Place type is a more general category than place name.

4. Country – (free text)
  - 4.1.RULE: Spell out full name, using American spelling; **common abbreviations are OK.**
  - 4.2.RULE: If given on ticket, enter; if verified from an external source, enter; if

unable to verify, leave blank. Acceptable external sources include: maps, encyclopedias, etc.

What happens to the database when the indexer does not follow a rule? Is that value irretrievable? Right. If the user doesn't know what variations to include in a search, the variations will not be retrieved. If the field is free text, we do not see what difference it makes whether the entry is USA or United States, or united states. Why burden the indexer with more input work than necessary? USA is definitely recognizable. If the user wants all tickets from the United States, this logic would make the user create the following search statement: us or usa or united states or united states of America or America – these would all be recognizable to someone reading the record, but places an unacceptable burden on someone trying to find all tickets from this country. Why not Czech Rep? We would want to enable the auto complete feature so that the indexer or the end user can enter "usa" and the program will change it to United States. Yes, an auto complete or substitution list can be quite helpful. Unless they have copies of the rules, end users won't know them anyway. It would be a foolish user who tried to search a controlled vocabulary without finding out what the controlled vocabulary contains.

#### 5. City – (free text)

5.1.RULE: Spell out full name, use American spelling

5.2.RULE: If given on ticket, enter; if verified from an external source, enter; if unable to verify, leave blank. Acceptable external sources include: maps, encyclopedias, etc.

**5.3 RULE: If in a foreign language that you understand, enter the English translation. Verify spelling in an external source.**

People in our beta test group were able to read foreign languages that we could not, which brought about another set of problems with language that we did not anticipate. If the indexer can translate the city from say Korean to English, we decided they should do so. This decision will lead to indexing inconsistency, and the searcher won't know whether to search for the Korean name or the English name. Three solutions are to enter both when known, or to enter only the one given on the ticket (which will usually be the functional equivalent of entering the city name in its own language), or to create a city name authority which shows both the native language name and the English translation. The third option is the best, but of course also the most expensive.

#### 6. Language – (free text)

6.1.RULE: Spell out full name with American spelling

**6.2 RULE: Enter for every language present; make each language a separate entry (F7)**

Our rule was so nonspecific that the beta team was left wondering how exactly to enter more than one language. They opted to use slashes (/) between languages when we wanted them to list each language separately.

#### 7. Currency value – (free text)

7.1.RULE: Enter original currency amount in numerals **with original decimal and comma notations** and the currency type **spelled in American English with the New York Stock Market equivalent abbreviations**. Ex. 50.00 HK dollars, 5,000 lira, 25.00 guilders, 35.00 Mex dollars, when you say “spelled in American English,” do you actually mean using English (Roman) characters (as opposed to Cyrillic, or Chinese, etc.)? I interpreted what you said literally when I read Part A, but perhaps that wasn’t what you meant.

7.2.RULE: Enter "free" if no fee is charged

The many different ways that countries around the world delineate their currency is a problem that we did not solve. Europeans use the decimal point and comma exactly opposite the American usage. We decided to enter the currency exactly as given on the ticket rather than translating everything to the American system. We decided to use the stock market equivalents for names of currencies, such as HK dollars, US dollar, Can dollar, Aus dollar. Mex dollar. It would be useful to provide your indexer and searchers with an authority list of the stock market equivalents.

8. Content –**Validation List**: map, hours of operation, serial #, history, address, phone

number, rules, **logo, block stamp**

8.1.RULE: Definition of history: historical information about the attraction, exhibit, point of interest, etc.

8.2 RULE: Definition of logo: **an insignia that might be near the company or place name**

8.3 RULE: **If content is in a foreign language that you understand, enter translated values for that content.**

8.4 RULE: **This field allows for multiple entries, if the object contains all, enter each value**

8.5 RULE: **Look on face and reverse sides of object**

This field gave our beta testers the most trouble because they misunderstood the concept and did not make use of the validation list. I don’t understand how not using the validation list was possible. If you had validated that field, then only the values in the validation list would be allowed by the software. Did you create a written authority list for them but neglect to create a validation list? Instead of entering the values given on the list to indicate the type of content present on the object, the beta testers entered the content itself. For example: they listed all the rules given on the object, gave the actual telephone number, hours of operation, etc. Why the validation list did not function automatically will be discussed in the next section, issues that came up. Oh, okay. One thing that might have helped prevent this situation with the Content field would have been to name the field differently. If we had called it “Content Type” it may have been more clear that we were intending the user to enter the elements that existed on the ticket instead of the actual text from the ticket itself. Yes, the names of fields are more important than you’d think initially. Also, you should say specifically that the text of the rules or historical information should not be entered; that the purpose is to indicate that there are rules or historical information on the document, not to provide the rules or history itself.

People in our beta test group were able to read foreign languages that we could not, which brought about another problem here as well. If the indexer can read content that is not in English they should translate it and include content values for it. This is particularly important because some languages use non-Roman characters, and will not be able to be entered into the database without translation. Unless you can hire indexers with the needed language expertise, it might be useful to have a value for this field that indicates that there is content in the local language – thus alerting your searchers, who may read the language of the country their researching, to the fact that the information may be present.

9. Visual image – **Validation List:** representative, decorative, none  
 9.1.RULE: Definition: representative : a photograph, drawing, or other image that represents it's not a good idea to use a word in its own definition. It would be better to say "that is a literal portrayal of" the actual attraction, and decorative: a design, graphic, or color element that decorates the ticket. and for decorative, you could say "that is not intended to depict the actual attraction"  
 9.2 RULE: This field allows for multiple entries.  
 9.3 RULE: Look on face and reverse sides of object.

This field also gave our beta testers much trouble because we did not clarify the definitions of representative and decorative nor give examples. We also needed more complete instructions.

#### 10. Logo—yes/no

Our beta testers convinced us that logo should be a value in "Content" rather than being a field of its own and we agreed. They found it hard to distinguish between a possible logo, a small graphic, and a block print (hanko), which are common insignias throughout the Far East.

10. Date – (free text)  
 11.1.RULE: 4-digit year/2-digit month/2-digit day (separated by slashes)  
~~11.2.RULE: Include time of day if listed, with am/pm if indicated~~  
~~11.3 RULE: Enter time as a separate entry on a second line (F7).~~

We decided that time of day did not indicate anything relevant. I'm inclined to agree with you.

11. Validated – **Validation List:** yes/no  
 12.1.RULE: Definition of validated: identifying stamp, punch, mark or tear to indicate whether the ticket is new or used

We decided to change the part of speech from noun to verb as this field signifies the act of validating the ticket, whereas validation was easily confused with Validation List, an entirely different concept. Oh, good point.

12. Physical Condition –**Validation List**: poor, fair, good, excellent  
 13.1.RULE: Definitions: poor – ripped, dirty; fair – folded, wrinkled, or bent; good – not ripped, folded, bent, or dirty; excellent – new, unused  
 13.3.RULE: Definition of ripped vs. torn: Ripped is unintentional damage or wear and tear; torn is a perforated section that has been removed to indicate that the ticket is no longer valid.

Your format of red changes with blue explanations makes this section very clear and easy to read.

### **Issues related to building data structures**

All in all, designing our database caused some interesting problems. We can now better understand a number of factors that have a direct effect on indexing quality. A database is a store of data about a selected part of the real world that is intended to be used for particular purpose. Indeed, developing a database involves several functions: planning and designing a data structure, planning units of analysis, creating entities and attributes, and selecting and formulating rules. The whole process of building the conceptual schema requires checking, rechecking, and modification of decisions. After all, the purpose of evaluation is to lead to improvements. Our findings are preliminary and tentative, but some interesting results did emerge.

Does the structure evolve from the data? In our case, yes. We examined the data and developed a structure (fields and their values) that would enable aggregating and segregating the data. Is this database, then, a self-enclosed unit frozen in time? No, at some point the data has to depart from the structure; or rather the structure has to depart from the data and exist independently of it. **This is an interesting way of looking at (and expressing) this idea.** The structure should allow for input of new data that might not adhere as closely to the existing values in the existing fields. There are ways to plan for this during the design phase. For instance in our “Place Type” field, we created a

validation list based on the objects in our current collection. But this limits the possible values to the ones we entered. We should have brainstormed all types of attractions and well-known places of interest, such as national parks, holy places, caves, zoos, volcanoes, dams, art galleries, opera houses, amusement parks, and put them all on the validation list, even before instances came up of having to use them. This principle governs Library of Congress subject headings, which categorize the entire universe of knowledge whether or not there are books currently under each subject heading. [Actually, it's more in line with the philosophy of the Dewey Decimal System than LC – historically LC's position was that their classification and subject headings were to be based on their existing collection, whereas Dewey felt it was important to create a structure that would fit everything.](#)

When creating the fields we first defined our users and then examined the objects from the standpoint of our users' needs. This forced us to move from the general to the specific when deciding on the attributes that we would develop into field. We realized that a too detailed, or minimalist approach, when examining physical characteristics, such as the approach we used with the postcard exercise in class, would lead us in the wrong direction. We had to think in a hierarchical, top down, fashion, to arrive at what later became the order of our fields. This order needs revising to maintain a truly top-down hierarchy. "Place type", the most abstract characteristic, should be #2. "Content," #8 and "Visual Image," #9 should also be moved up in the hierarchy, as they are more abstract than "Place name," #2, "Country," #4, "City," #5, "Language," #6, or "Currency," #7.

We had one unexpected event that arose in the last minute. When we received our database from the beta team, we found out that they did not use our copy of the

DBTextWorks structure, which we had given to them on a floppy disk. Because the laptop they used did not have a floppy disk drive, the beta team member responsible reconstructed her own version of our database on DBTextWorks, without telling her team members. Both teams discovered this when she did not have a disk copy or print out of the records to hand us. Later she e-mailed us her copy of the records from her version of the database. This explains some very fundamental differences in the two fields. Since she did not enter any validation lists, she was able to enter free text values in fields we had designed as limited by validation lists. For example, in the “Content” field she entered the actual content on the tickets, instead of our categories of content, the values given on the validation list. [Ah. Now I understand.](#)

Nevertheless, putting our database through a beta testing process was very helpful in illustrating the pitfalls of creating a stable database with a clear rule set. We have learned that we needed to spend more time clearly describing the fields and how we intended for them to function; even when we thought our rules were clear, there turned out to be additional room for misinterpretation. In order to ensure consistent records, the rules need to be incredibly detailed. We have also learned that we needed to define our terms more completely and more often than we had expected; while what we meant by certain words was understood within our group, the beta group struggled to determine what was intended. A clearer description and rationale for the database might have also helped with these definition problems. Finally, going through the beta testing process illustrated that the margin for error was larger than we had realized. If we were to do the project again, we would spend more time clarifying out rules and definitions to make for a more consistent and therefore more useful end product.

A very insightful discussion. Excellent work throughout the project.

**345**