Introduction to Informatics

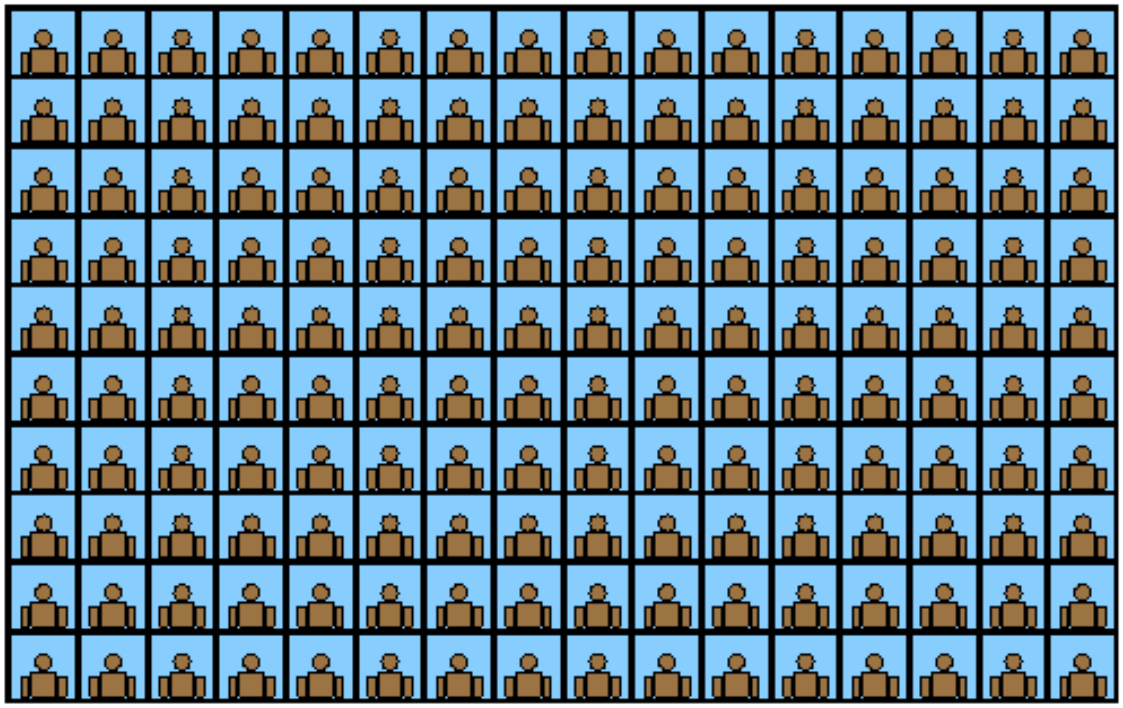
Lecture 26:

Information Technology in the Real World

Databases

Luis M.Rocha and Santiago Schnell

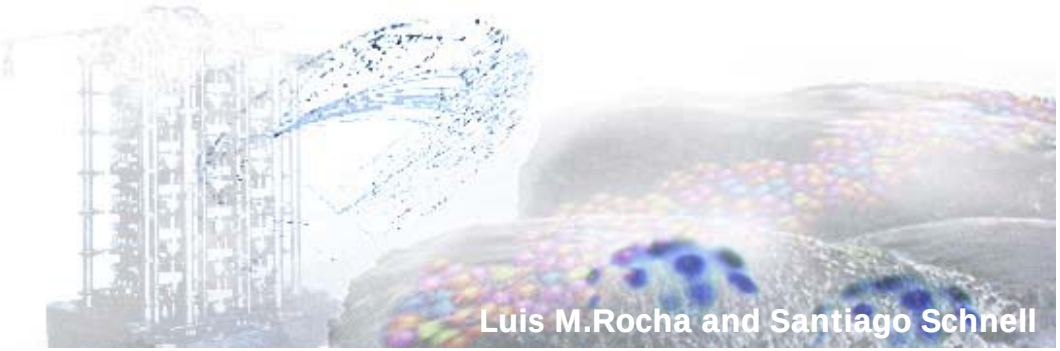
NO MORE LABS !!!



Luis M.Rocha and Santiago Schnell

Exam Schedule

- 11595
  - Midterm
    - March 1st (Thursday)
      - Regular Class time
  - Final Exam
    - May 3rd (Thursday)
      - 7:15-9:15 p.m.



Readings until now

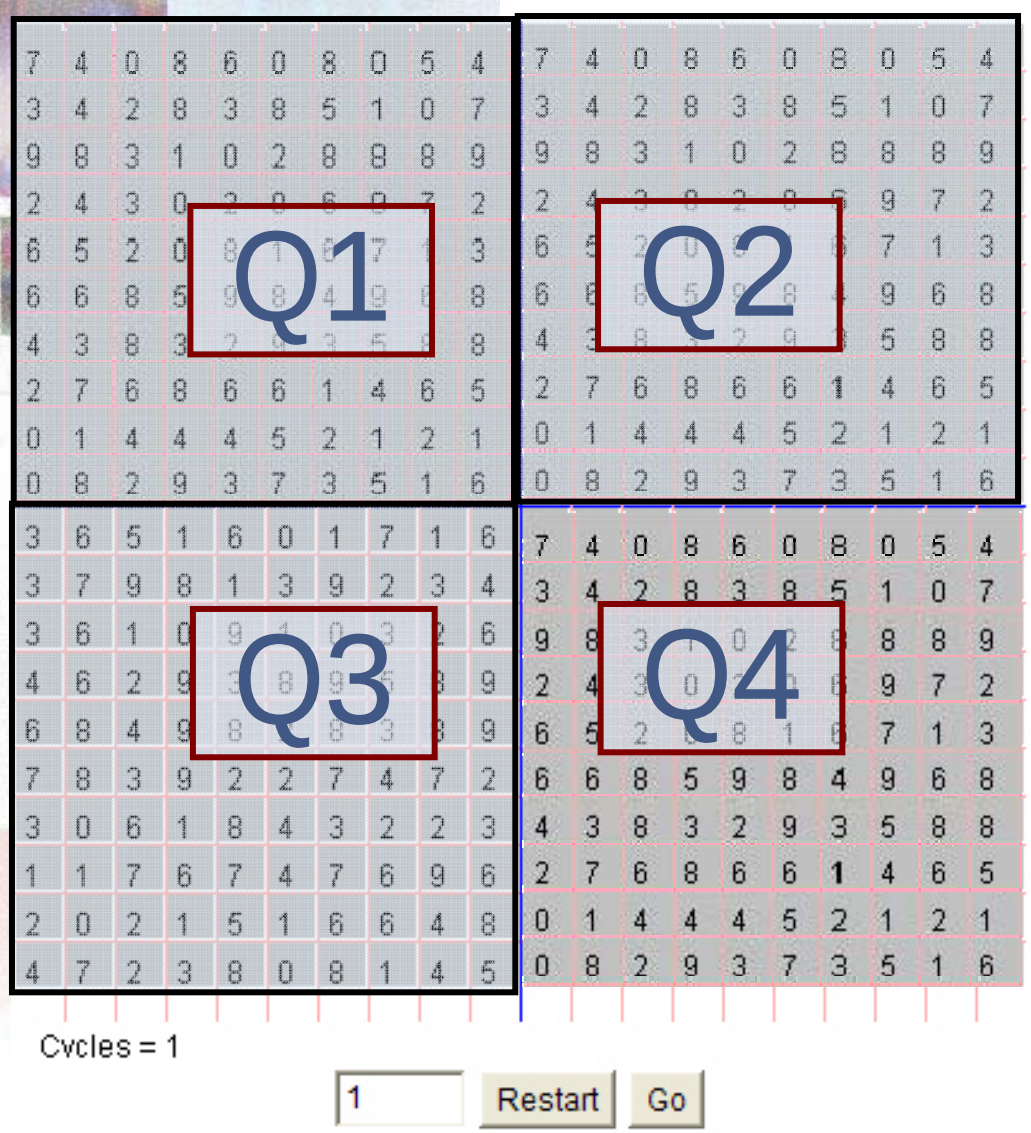
- Lecture notes
  - Posted online
    - http://informatics.indiana.edu/rocha/i101
      - *The Nature of Information*
      - *Technology*
      - *Modeling the World*
  - @ *infoport*
    - *http://infoport.blogspot.com*
  - From course package
    - Von Baeyer, H.C. [2004]. *Information: The New Language of Science*. Harvard University Press.
      - Chapters 1, 4 (pages 1-12)
      - Chapter 10 (pages 13-17)
    - From Andy Clark's book "*Natural-Born Cyborgs*"
      - Chapters 2 and 6 (pages 19 - 67)
    - From Irv Englander's book "*The Architecture of Computer Hardware and Systems Software*"
      - Chapter 3: Data Formats (pp. 70-86)
    - Klir, J.G., U. St. Clair, and B.Yuan [1997]. Fuzzy Set Theory: foundations and Applications. Prentice Hall
      - Chapter 2: Classical Logic (pp. 87-97)
      - Chapter 3: Classical Set Theory (pp. 98-103)
    - Norman, G.R. and D.L. Streinrt [2000]. *Biostatistics: The Bare Essentials*.
      - Chapters 1-3 (pages 105-129)
      - OPTIONAL: Chapter 4 (pages 131-136)
      - Chapter 13 (pages 147-155)
      - Chapter 5 (pages 141-144)
    - Igor Aleksander, "Understanding Information Bit by Bit"
      - Pages 157-166
    - Ellen Ullman, "Dining with Robots"
      - Pages 167-172

Assignment Situation

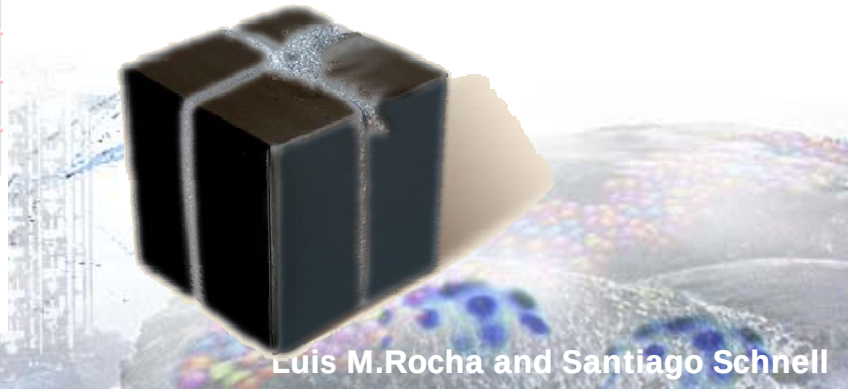
- Labs
  - Past
    - Lab 1: Blogs
      - Closed (Friday, January 19): Grades Posted
    - Lab 2: Basic HTML
      - Closed (Wednesday, January 31): Grades Posted
    - Lab 3: Advanced HTML: Cascading Style Sheets
      - Closed (Friday, February 2): Grades Posted
    - Lab 4: More HTML and CSS
      - Closed (Friday, February 9): Grades Posted
    - Lab 5: Introduction to Operating Systems: Unix
      - Closed (Friday, February 16): Grades Posted
    - Lab 6: More Unix and FTP
      - Closed (Friday, February 23): Grades Posted
    - Lab 7: Logic Gates
      - Closed (Friday, March 9): Grades Posted
    - Lab 8: Intro to Statistical Analysis using Excel
      - Closed (Friday, March 30): Grades Posted
    - Lab 9: Data analysis with Excel (linear regression)
      - Closed (Friday, April 6): Grades Posted
    - Lab 10: Simple programming in Excel and Measuring Uncertainty
      - April 12 and 13, Due April 20

- Assignments
  - Individual
    - First installment
      - Closed: February 9: Grades Posted
    - Second Installment
      - Past: March 2: Grades Posted
    - Third installment
      - Past: Grades Posted
    - Fourth Installment
      - Presented April 10th, Due April 20th
  - Group
    - First Installment
      - Past: March 9th, graded
    - Second Installment
      - Past: April 6th Graded
    - Third Installment
      - Presented Thursday, April 12; Due Friday, April 27

Individual Assignment – Part IV

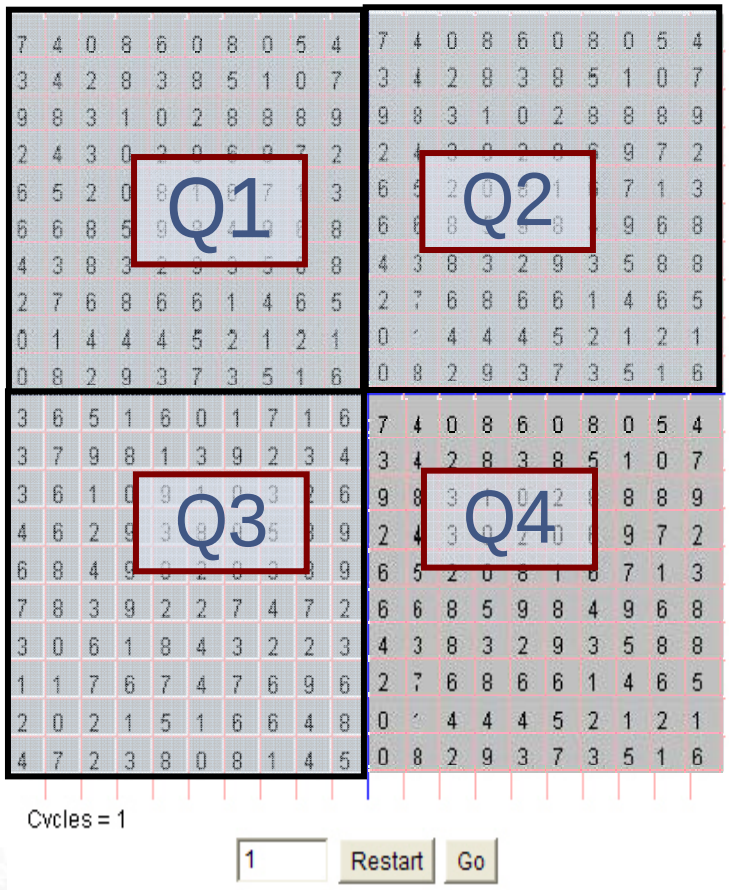


- Step by step analysis of “dying” squares
  - 4th Installment
    - Presented: April 10th
    - Due: April 20th
- Use inductive and deductive reasoning
  - To uncover the algorithm in each quadrant
    - Build from inductive knowledge accumulated so far



Summary of Black Box

- Quadrant 1
  - At the random initial state
    - All numbers have equal probability of being initially present
    - But the probability of changes are different
- In Any State
  - Any number changes depending on its neighbors
  - It ‘gravitates’ towards the smallest number that it ‘sees’ most often.
  - Odd and Even numbers do not show different behavior
- What is the Algorithm?



Summary of Black Box

- Quadrant 3
  - At the random initial state
    - All numbers have equal probability of being initially present
    - But the probability of changes are different
- In Any State
  - 0 can only change to 0
  - 5 can only change to 5 or 0
  - Even digits always change to even digits
  - Odd digits could change to any other digit
- What is the Algorithm?

	n(i)	p(i)
0	27	0.27
1	4	0.04
2	12	0.12
3	4	0.04
4	12	0.12
5	9	0.09
6	12	0.12
7	4	0.04
8	12	0.12
9	4	0.04

1. $0 \rightarrow 0$
2. $\{5\} \rightarrow \{0, 5\}$
3. $\{2, 4, 6, 8\} \rightarrow \{0, 2, 4, 6, 8\}$
4. $\{1, 3, 7, 9\} \rightarrow$ $\{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$

Summary of Black Box

- Quadrant 2
  - At the random initial state
    - All numbers have equal probability of being initially present
    - But the probability of changes are different
- In Any State
  - 0 can only change to 0
  - 5 can only change to 5 or 0
  - Even digits always change to even digits
  - Odd digits could change to any other digit
- What is the Algorithm?

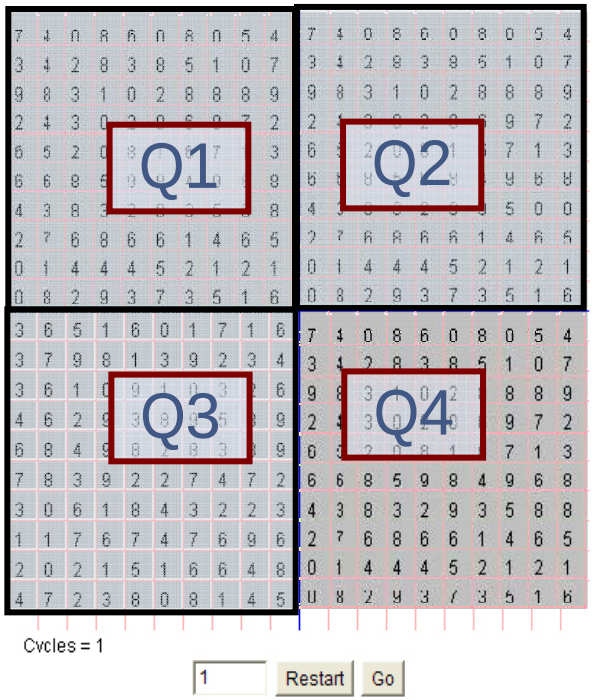
1. $0 \rightarrow 0$
2. $\{5\} \rightarrow \{0, 5\}$
3. $\{2, 4, 6, 8\} \rightarrow \{0, 2, 4, 6, 8\}$
4. $\{1, 3, 7, 9\} \rightarrow$ $\{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$

Possible Operations Q2 and Q3

Operator	Meaning	Excel	Example
()	Brackets, grouping	()	y = (a + b) * (c + d)
*	Multiplication	*	i=j*k
+	Add	+	i = i+1
-	Subtract	-	i=j-3.2
/	Real division	/	i=8/5 = 1.6
div	Integer division	Quotient (a,b)	i=8/5 = 1
Mod, %	remainder	Mod (a, b)	i=8 mod 5 = 3
ROUND	Rounds	ROUND (a, d)	i=ROUND(3.67,0) = 4
INT	Integer Part	INT	i=INT(3.67) = 3
rand	Random number	Rand() RandBetween(a,b)	i=rand(n)

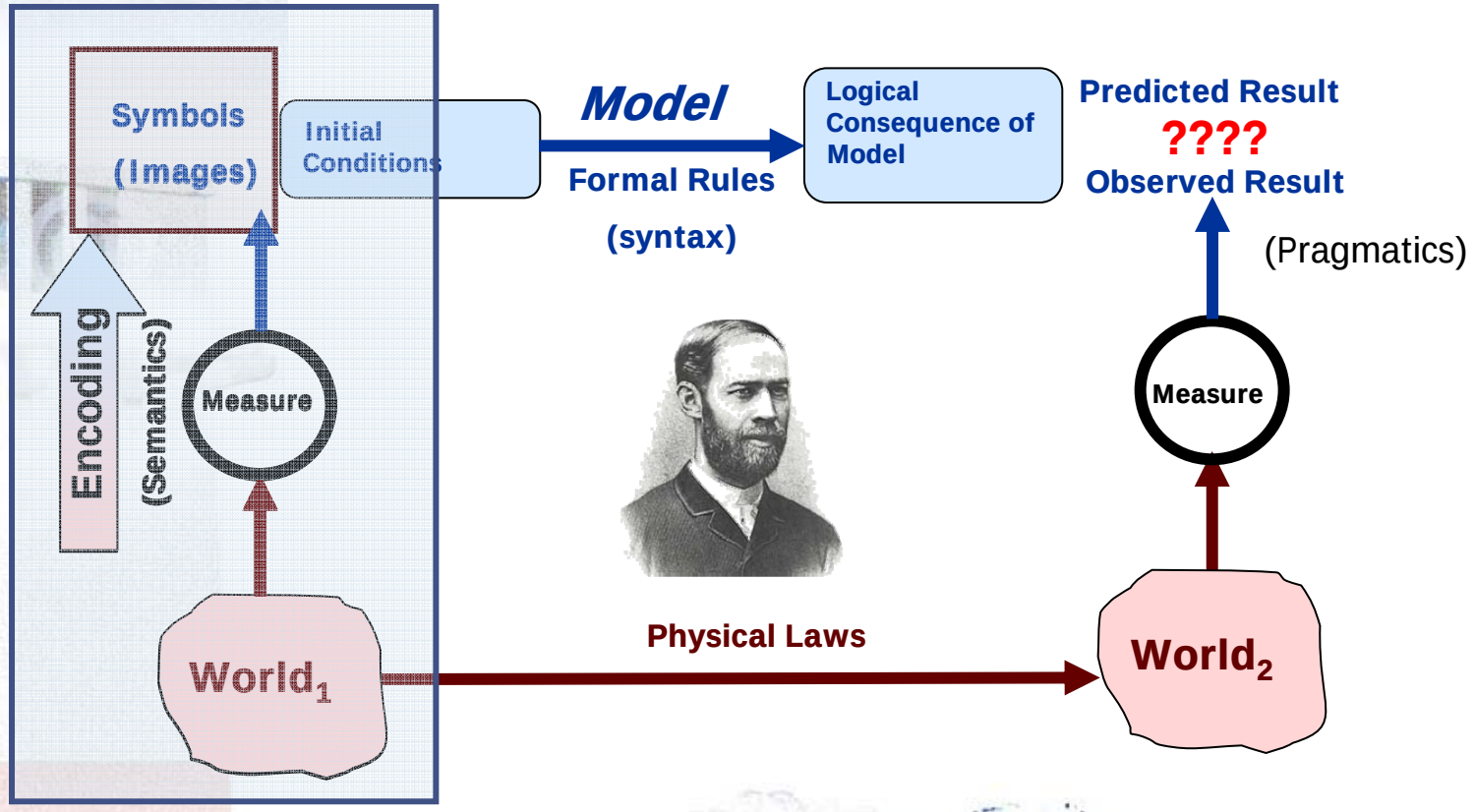
Tip for Individual Assignment

- Quadrant Q
  - There are 100 cells in each 10x10 quadrant
    - C = 1...100
  - Each cell can take one of 10 colors
    - V(C)=0..9
      - is the value of the cell
      - This is the state cell C is in
  - Random initialization of quadrant Q at cycle 1
    - For c=1 to 100 do
    - V(C) ← randbetween(0,9) {random number 0 to 9}
    - EndFor
    - Cycle ← 1
  - Run for Number of cycles
    - n ← Input dialog
    - For k=1 to n do
    - Cycle ← cycle+1
    - {Pick random cell}
    - C ← randbetween(1,100)
    - {Update the value of the cell (NOT THE REAL THING)}
    - V(C) ← ((V(C) * randbetween(0,9)) div 2) – 5*x
    - EndFor
  - X may be a hidden variable
    - X ← ???



The Modeling Relation

Hertz' Modeling Paradigm

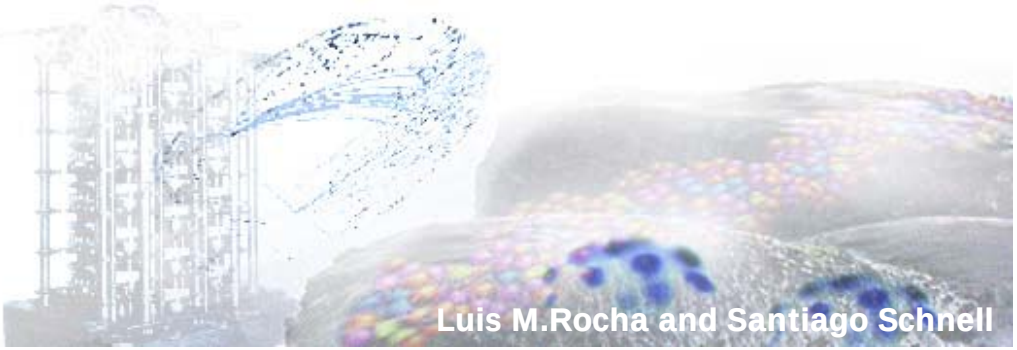


- Organizing Data
  - After encoding
  - Modern Problems Require large storage capabilities

The Entity-Relationship Model

- Conceptual ***Data Model***
  - A kind of "pseudocode" for ***models of data storage***
- What should we consider?
  - What are the interesting entities and relationships in our model of reality?
  - What information about these entities and relationships do we need to store?
  - What are the reality constraints and rules that must hold?

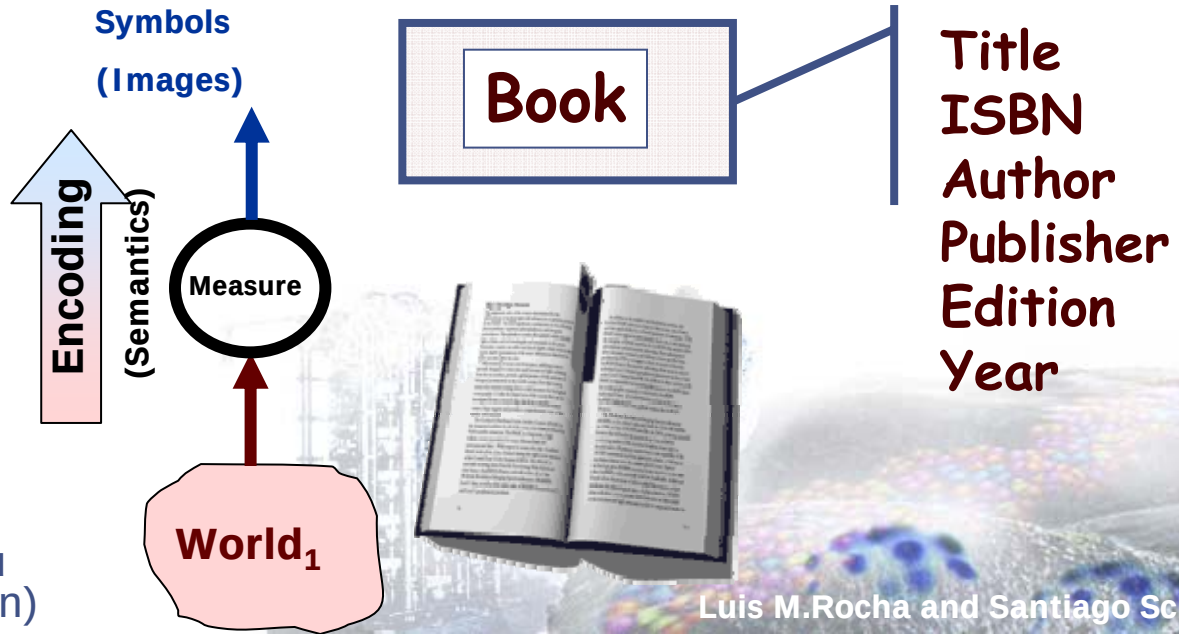
Peter Chen
(1976)



Adapted from Yuqing Melanie Wu
(I308: Information Representation)

Entities in Data Modeling

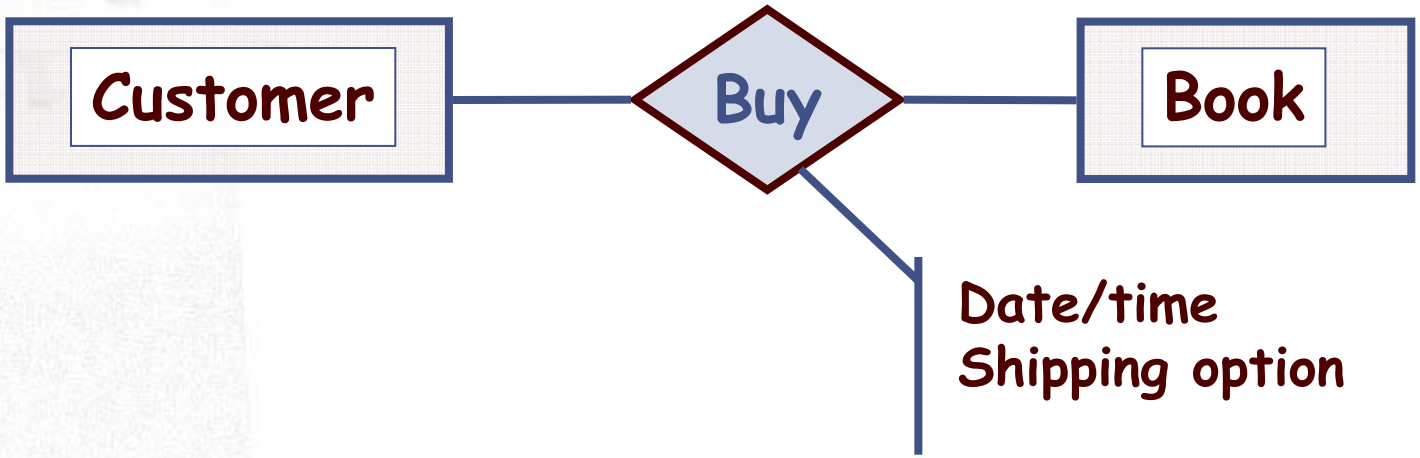
- Objects, people, places
  - Basically ***a noun*** : a discrete object
    - Choose a meaningful name
  - Represented by a rectangle
  - Attributes
    - Describe the proprieties of an entity



Adapted from Yuqing Melanie Wu
(I308: Information Representation)

Relationships in Data Modeling

- Relationship:
  - An association among two or more entities.
  - Verbs



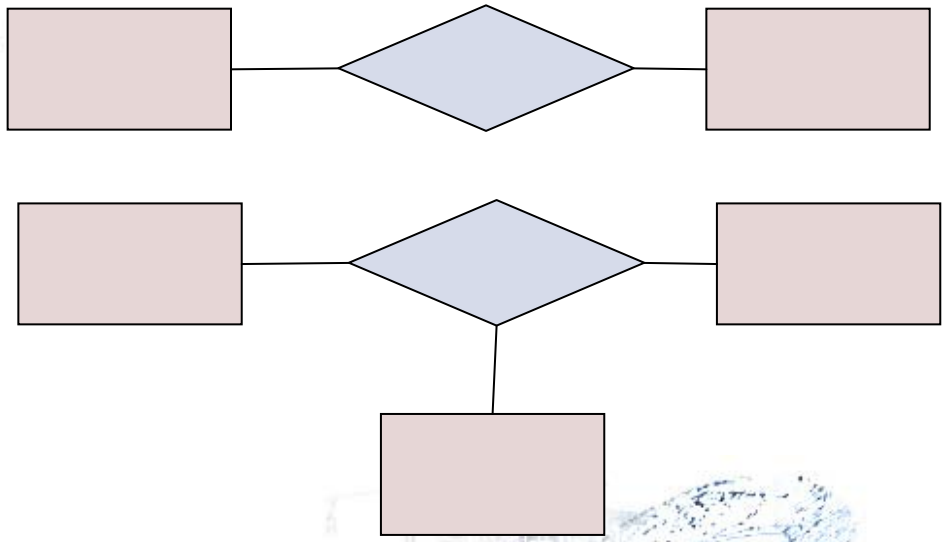
- Attributes also describe relationships

Adapted from Yuqing Melanie Wu
(I308: Information Representation)

Luis M.Rocha and Santiago Schnell

Arity of Relationships

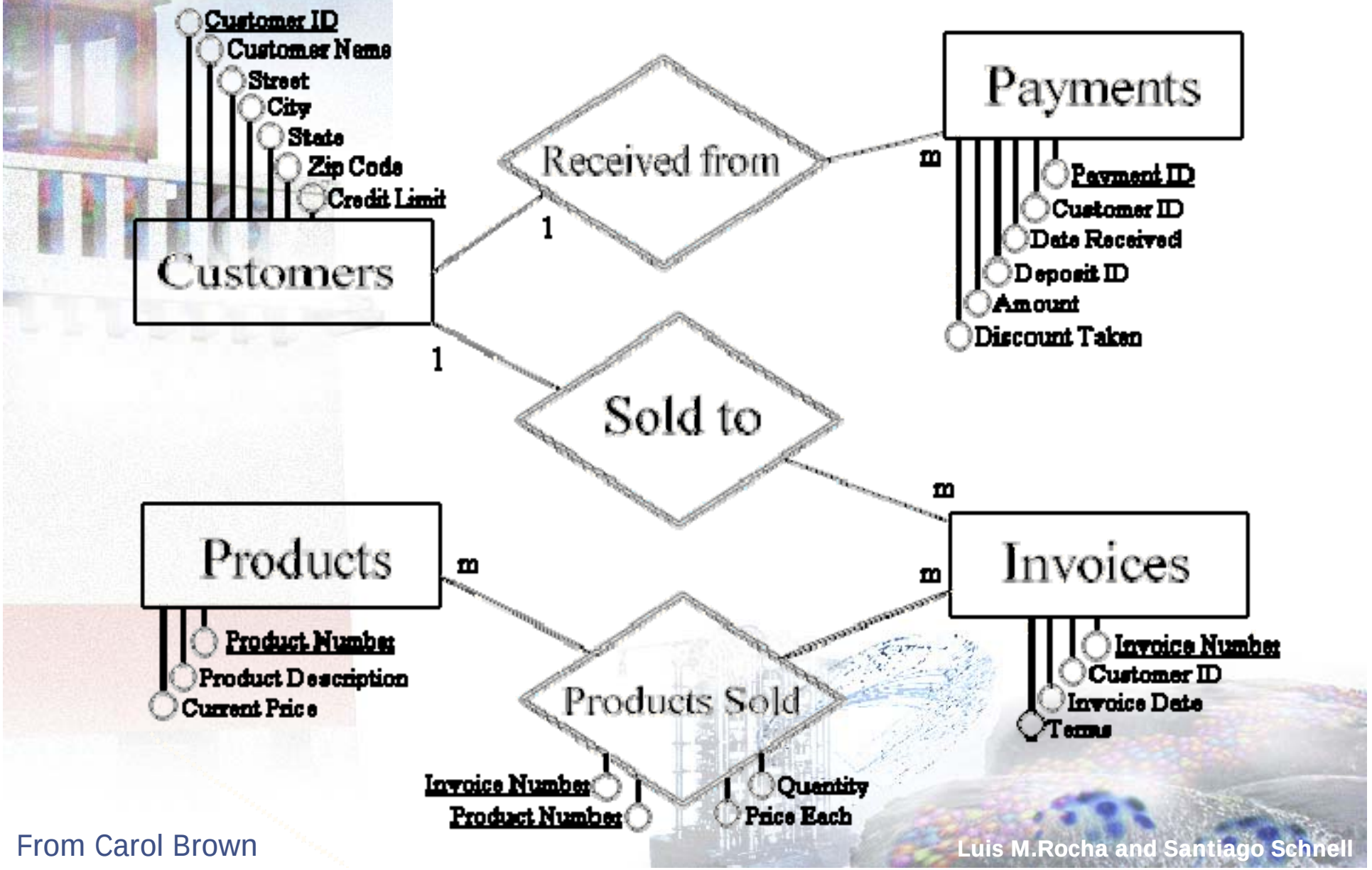
- The number of entities participate in a relationship
  - Binary, ternary, N-ary



Adapted from Yuqing Melanie Wu
(I308: Information Representation)

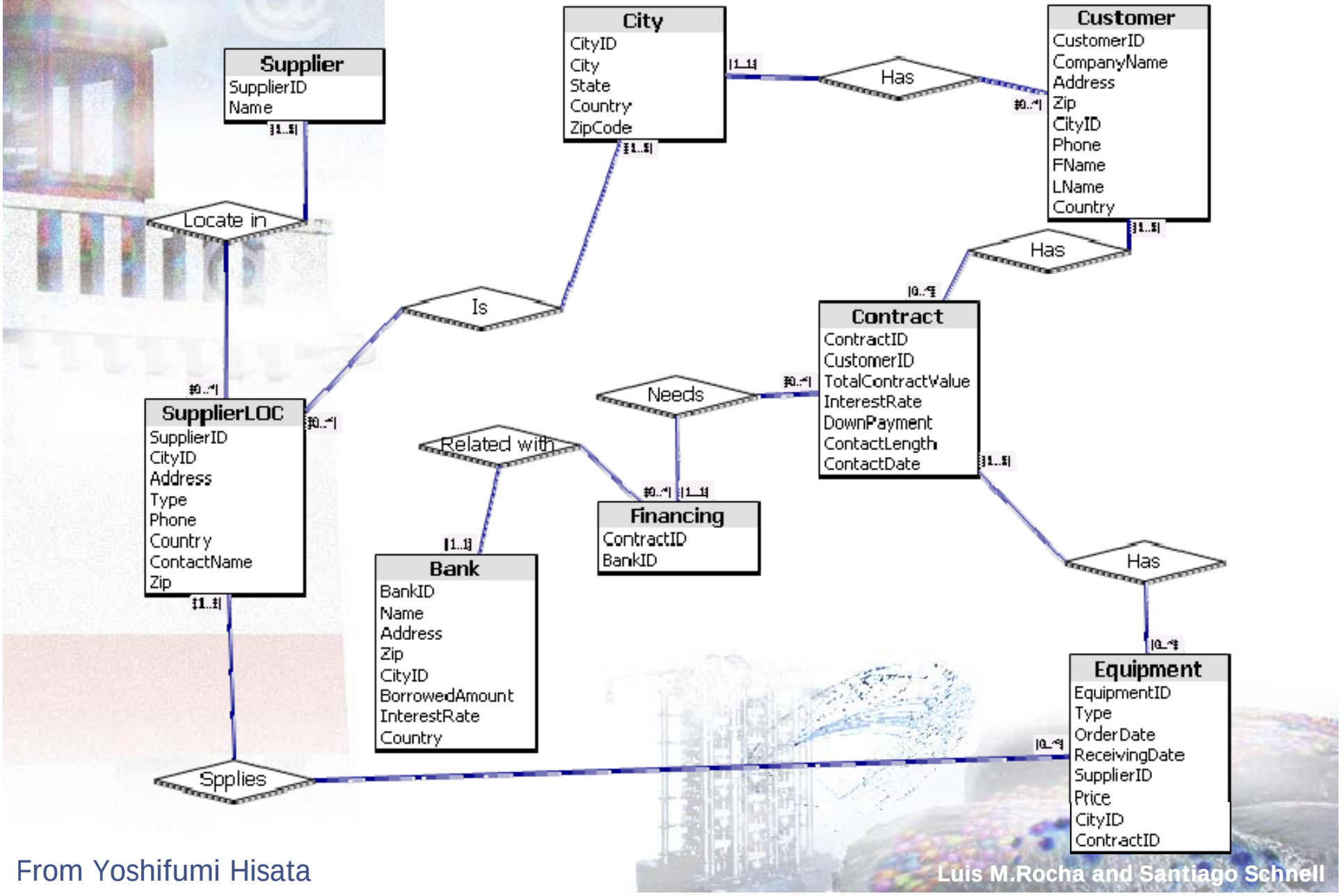
Luis M.Rocha and Santiago Schnell

ER Data Modeling Example



From Carol Brown

ER Data Modeling Example 2

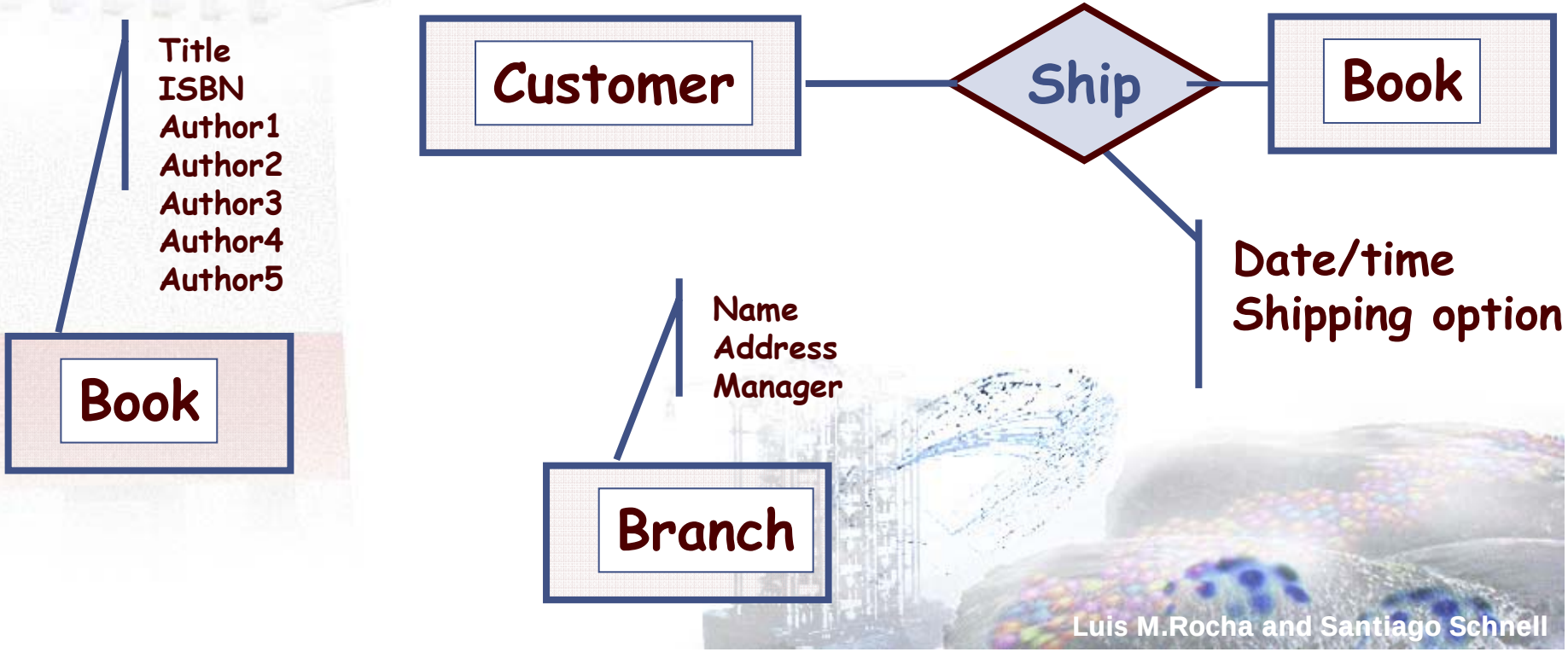


From Yoshifumi Hisata

Luis M.Rocha and Santiago Schnell

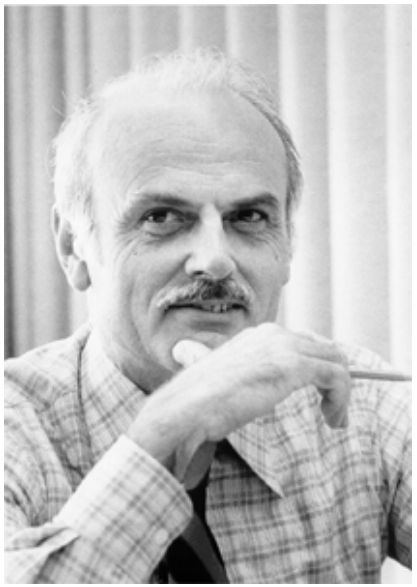
Try this at home…

- How to represent the following?
  - A book can have no more than 5 authors
  - A customer has to specify the shipping option
  - Each branch has only one manager.



The Relational Database Model

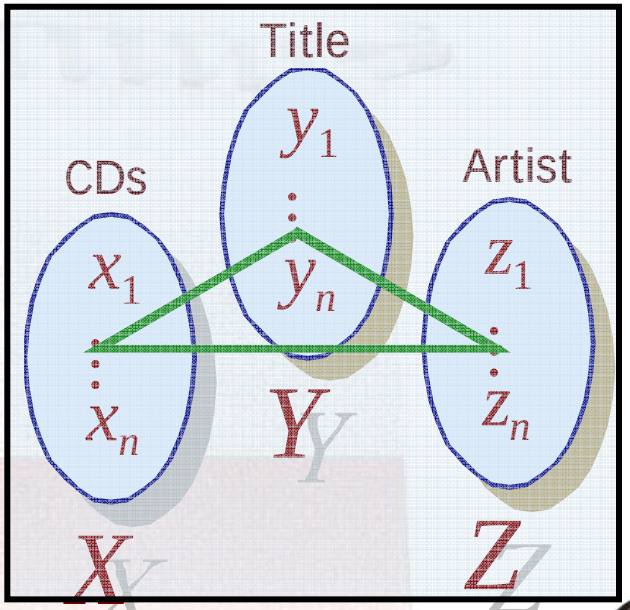
- Relational database management system (RDBMS)
  - Most popular commercial database type.
  - a data model based on ***logic*** and ***set theory***.
- invented by Ted Codd in 1970
  - Oxford, IBM, U. Michigan, IBM
- System R
  - IBM's San Jose research center
  - Structured English Query Language ("SEQUEL")
    - Data Manipulation Language (DML)
  - SEQUEL was later condensed to SQL due to a trademark dispute
  - In 1979, Relational Software, Inc. (now Oracle Corporation) introduced the first commercially available implementation of SQL

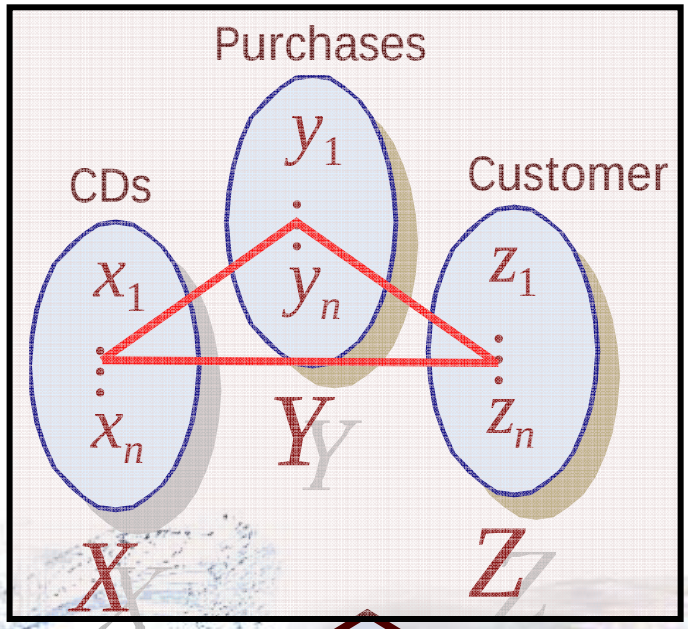


Ted Codd

The Relational Model

- All data are represented as mathematical relations
  - Represent the presence of association, interaction or interconnectedness between the elements of two or more sets.
  - A relation associates the elements of 2 or more sets
    - Set of books with sets of attributes (*entities*)
    - Set of purchases with sets of attributes (*relationships*)
  - Tables store relations



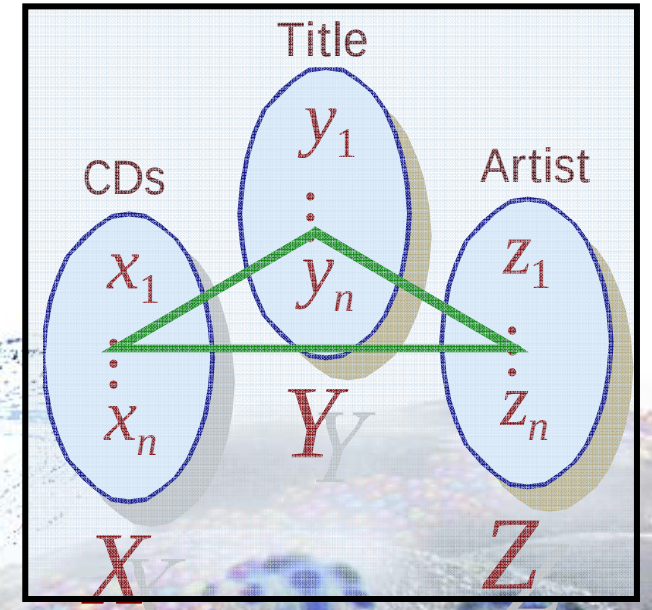


The Relational Database Model

- A relational database is a collection of tables
  - 2-dimensional
- Each table has a unique name in the database.
- Tables define Relations
  - Columns (number of sets)
    - Attributes plus key (*primary set*)
  - Row (number of relation instances)
    - A table is a set of rows: tuples

CDs

ID	Title	Artist
3592	Yes I am a Witch	Yoko Ono
2678	Big	Macy Gray
0623	Sound of Silver	LCD Soundsystem
0321	Welcome to Planet Sexor	Tiga
8854	Transparent Things	Fujiya & Miyagi



Example

Attributes (columns)

Relation (table)

Customer

Phone	Name	Address
812-123-4567	Tom	408 3rd st. Bloomington, IN
812-304-2378	Bill	#113, Redbud Hall, Bloomington, IN
812-856-1190	Kate	1205, Maritime ct. Bloomington, IN
812-754-9567	Mary	#901 10th St. Bloomington, IN
317-897-4536	Pam	2400 Rd135, Greenwood, IN
812-906-2486	Jeff	#208 Union Ave. Bloomington, IN

Tuples (rows)

Degree = 3

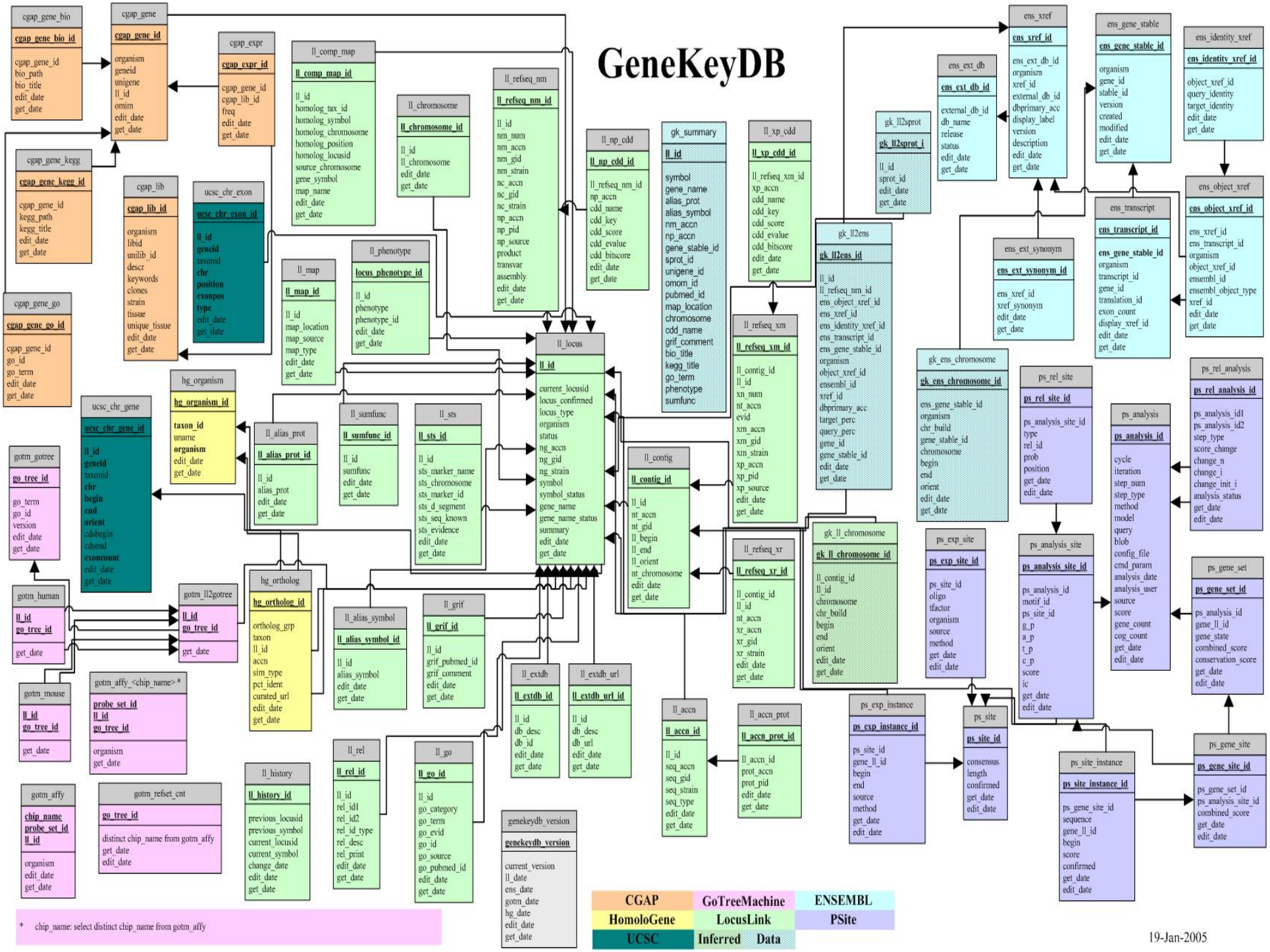
Cardinality = 6

From Yuqing Melanie Wu (I308: Information Representation)

Schema and Instance

- Database schema
  - Metadata or Model
  - The logical design of a database
    - E.g. using the *entity-relationship model*
      - *Entity → Table*
      - *Attribute → Columns*
      - *Relationship → Table*
  - Specifies names of tables/relations (*entities and relationships*), plus names and types of each column (*attributes*)
- Database instance
  - A snapshot of the data in the database at a given instant in time.

Adapted from Yuqing Melanie Wu
(I308: Information Representation)


GeneKeyDB
cgap_gene_bio
cgap_gene_bio_id
cgap_gene_id
bio_path
bio_title
edit_date
get_date
cgap_gene
cgap_gene_id
organism
geneid
unigene
ll_id
omim
edit_date
get_date
cgap_expr
cgap_expr_id
cgap_gene_id
cgap_lib_id
freq
edit_date
get_date
ll_comp_map
ll_comp_map_id
ll_id
homolog_tax_id
homolog_symbol
homolog_chromosome
homolog_position
homolog_locusid
source_chromosome
gene_symbol
map_name
edit_date
get_date
ll_chromosome
ll_chromosome_id
ll_id
ll_chromosome
edit_date
get_date
ll_refseq_nm
ll_refseq_nm_id
ll_id
nm_num
nm_accn
nm_gid
nm_strain
nc_accn
nc_gid
nc_strain
np_accn
np_pid
np_source
product
transvar
assembly
edit_date
get_date
ll_np_cdd
ll_np_cdd_id
ll_refseq_nm_id
np_accn
cdd_name
cdd_key
cdd_score
cdd_evalue
cdd_bitscore
edit_date
get_date
gk_summary
ll_id
symbol
gene_name
alias_prot
alias_symbol
nm_accn
np_accn
gene_stable_id
sprot_id
unigene_id
omom_id
pubmed_id
map_location
chromosome
cdd_name
grif_comment
bio_title
kegg_title
go_term
phenotype
sumfunc
ll_xp_cdd
ll_xp_cdd_id
ll_refseq_xm_id
xp_accn
cdd_name
cdd_key
cdd_score
cdd_evalue
cdd_bitscore
edit_date
get_date
gk_ll2sprot
gk_ll2sprot_i
ll_id
sprot_id
edit_date
get_date
ens_ext_db
ens_ext_db_id
external_db_id
db_name
release
status
edit_date
get_date
ens_xref
ens_xref_id
ens_ext_db_id
organism
xref_id
external_db_id
dbprimary_acc
display_label
version
description
edit_date
get_date
ens_gene_stable
ens_gene_stable_id
organism
gene_id
stable_id
version
created
modified
edit_date
get_date
ens_identity_xref
ens_identity_xref_id
object_xref_id
query_identity
target_identity
edit_date
get_date
ens_object_xref
ens_object_xref_id
ens_xref_id
ens_transcript_id
organism
object_xref_id
ensembl_id
ensembl_object_type
xref_id
edit_date
get_date
ens_transcript
ens_transcript_id
ens_gene_stable_id
organism
transcript_id
gene_id
translation_id
exon_count
display_xref_id
edit_date
get_date
ens_ext_synonym
ens_ext_synonym_id
ens_xref_id
xref_synonym
edit_date
get_date
cgap_gene_kegg
cgap_gene_kegg_id
cgap_gene_id
kegg_path
kegg_title
edit_date
get_date
cgap_lib
cgap_lib_id
organism
libid
unilib_id
descr
keywords
clones
strain
tissue
unique_tissue
edit_date
get_date
ucsc_chr_exon
ucsc_chr_exon_id
ll_id
geneid
taxonid
chr
position
exonpos
type
edit_date
get_date
ll_phenotype
locus_phenotype_id
ll_id
phenotype
phenotype_id
edit_date
get_date
ll_map
ll_map_id
ll_id
map_location
map_source
map_type
edit_date
get_date
gk_ll2ens
gk_ll2ens_id
ll_id
ll_refseq_nm_id
ens_object_xref_id
ens_xref_id
ens_identity_xref_id
ens_transcript_id
ens_gene_stable_id
organism
object_xref_id
ensembl_id
xref_id
dbprimary_acc
target_perc
query_perc
gene_id
gene_stable_id
edit_date
get_date
cgap_gene_go
cgap_gene_go_id
cgap_gene_id
go_id
go_term
edit_date
get_date
hg_organism
hg_organism_id
taxon_id
uname
organism
edit_date
get_date
ll_locus
ll_id
current_locusid
locus_confirmed
locus_type
organism
status
ng_accn
ng_gid
ng_strain
symbol
symbol_status
gene_name
gene_name_status
summary
edit_date
get_date
ll_refseq_xm
ll_refseq_xm_id
ll_contig_id
ll_id
xn_num
nt_accn
evid
xm_accn
xm_gid
xm_strain
xp_accn
xp_pid
xp_source
edit_date
get_date
gk_ens_chromosome
gk_ens_chromosome_id
ens_gene_stable_id
organism
chr_build
gene_stable_id
chromosome
begin
end
orient
edit_date
get_date
ps_rel_site
ps_rel_site_id
ps_analysis_site_id
type
rel_id
prob
position
get_date
edit_date
ps_rel_analysis
ps_rel_analysis_id
ps_analysis_id1
ps_analysis_id2
step_type
score_change
change_n
change_i
change_init_i
analysis_status
get_date
edit_date
ucsc_chr_gene
ucsc_chr_gene_id
ll_id
geneid
taxonid
chr
begin
end
orient
cdsbegin
cdsend
exoncount
edit_date
get_date
ll_alias_prot
ll_alias_prot_id
ll_id
alias_prot
edit_date
get_date
ll_sumfunc
ll_sumfune_id
ll_id
sumfunc
edit_date
get_date
ll_sts
ll_sts_id
ll_id
sts_marker_name
sts_chromosome
sts_marker_id
sts_d_segment
sts_seq_known
sts_evidence
edit_date
get_date
ll_contig
ll_contig_id
ll_id
nt_accn
nt_gid
ll_begin
ll_end
ll_orient
nt_chromosome
edit_date
get_date
gk_ll_chromosome
gk_ll_chromosome_id
ll_contig_id
ll_id
chromosome
chr_build
begin
end
orient
edit_date
get_date
ps_exp_site
ps_exp_site_id
ps_site_id
oligo
tfactor
organism
source
method
get_date
edit_date
ps_analysis
ps_analysis_id
cycle
iteration
step_num
step_type
method
model
query
blob
config_file
cmd_param
analysis_date
analysis_user
source
score
gene_count
cog_count
get_date
edit_date
ps_analysis_site
ps_analysis_site_id
ps_analysis_id
motif_id
ps_site_id
g_p
a_p
t_p
c_p
score
ic
get_date
edit_date
ps_gene_set
ps_gene_set_id
ps_analysis_id
gene_ll_id
gene_state
combined_score
conservation_score
get_date
edit_date
gotm_gotree
go_tree_id
go_term
go_id
version
edit_date
get_date
gotm_human
ll_id
go_tree_id
get_date
gotm_ll2gotree
ll_id
go_tree_id
get_date
hg_ortholog
hg_ortholog_id
ortholog_grp
taxon
ll_id
accn
sim_type
pct_ident
curated_url
edit_date
get_date
ll_alias_symbol
ll_alias_symbol_id
ll_id
alias_symbol
edit_date
get_date
ll_grif
ll_grif_id
ll_id
grif_pubmed_id
grif_comment
edit_date
get_date
ll_refseq_xr
ll_refseq_xr_id
ll_contig_id
ll_id
nt_accn
xr_accn
xr_gid
xr_strain
edit_date
get_date
ll_extdb
ll_extdb_id
ll_id
db_desc
db_id
edit_date
get_date
ll_extdb_url
ll_extdb_url_id
ll_id
db_desc
db_url
edit_date
get_date
ll_accn
ll_accn_id
ll_id
seq_accn
seq_gid
seq_strain
seq_type
edit_date
get_date
ll_accn_prot
ll_accn_prot_id
ll_accn_id
prot_accn
prot_pid
edit_date
get_date
ps_exp_instance
ps_exp_instance_id
ps_site_id
gene_ll_id
begin
end
source
method
get_date
edit_date
ps_site
ps_site_id
consensus
length
confirmed
get_date
edit_date
ps_site_instance
ps_site_instance_id
ps_gene_site_id
sequence
gene_ll_id
begin
score
confirmed
get_date
edit_date
ps_gene_site
ps_gene_site_id
ps_gene_set_id
ps_analysis_site_id
combined_score
get_date
edit_date
gotm_mouse
ll_id
go_tree_id
get_date
gotm_affy_<chip_name>*
probe_set_id
ll_id
go_tree_id
organism
get_date
gotm_affy
chip_name
probe_set_id
ll_id
organism
edit_date
get_date
gotm_refset_cnt
go_tree_id
distinct chip_name from gotm_affy
get_date
edit_date
ll_history
ll_history_id
previous_locusid
previous_symbol
current_locusid
current_symbol
change_date
edit_date
get_date
ll_rel
ll_rel_id
ll_id
rel_id1
rel_id2
rel_id_type
rel_desc
rel_print
edit_date
get_date
ll_go
ll_go_id
ll_id
go_category
go_term
go_evid
go_id
go_source
go_pubmed_id
edit_date
get_date
genekeydb_version
genekeydb_version
current_version
ll_date
ens_date
gotm_date
hg_date
edit_date
get_date
* chip_name: select distinct chip_name from gotm_affy
CGAP
GoTreeMachine
ENSEMBL
HomoloGene
LocusLink
PSite
UCSC
Inferred Data
19-Jan-2005

Primary Key

Key

Customer

Phone	Name	Address
812-123-4567	Tom	408 3rd st. Bloomington, IN
812-304-2378	Bill	#113, Redbud Hall, Bloomington, IN
……	……	……

Customer(Phone, Name, Address)

- The identifying labels for the elements of the primary set of a table
- Every instance (row) in the database must have a distinct primary key
- Every instance in the database must have a particular (non-null) value for the primary key.

Adapted from Yuqing Melanie Wu
(I308: Information Representation)

Book

Key

ISBN	Title	Publisher
12345	Java	MIT press
49082	Snow White	……
72936	Honeymoon	……

Customer

Key

Phone	Name	Address
812-123-4567	Tom	……
812-304-2378	Bill	……
812-856-1190	Kate	……

Book(ISBN, Title, Publisher)

Customer(Phone, Name, Address)

Sale

Key

ISBN	Phone	Price	Time
12345	812-123-4567	$20	Feb 2, 05
49082	812-123-4567	$25	Dec 20, 04
12345	812-856-1190	$19	……

Sale (ISBN, Phone,time,price)

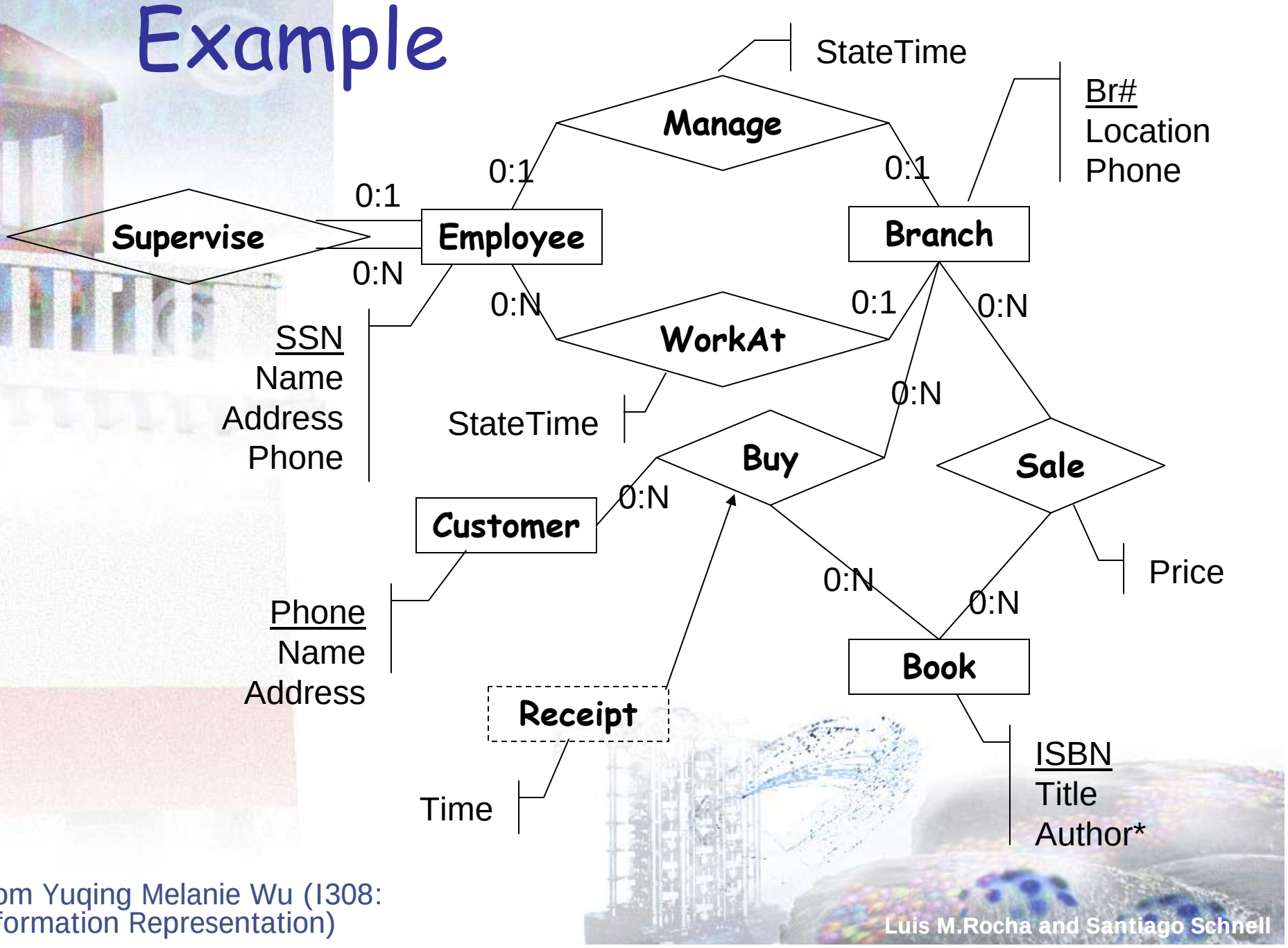
Come from primary keys in entities

Primary keys in Relationships

Adapted from Yuqing Melanie Wu
(I308: Information Representation)

Luis M.Rocha and Santiago Schnell

Example



From Yuqing Melanie Wu (I308: Information Representation)

Structured Query Language (SQL)

- The most popular computer language used to create, modify and retrieve data from relational database management systems. (Wikipedia)
  - Three subsets of SQL
    - Data Definition Language (DDL)
    - Data Manipulation Language (DML)
    - Data Control Language (DCL) (for authorization)

Data Definition Language

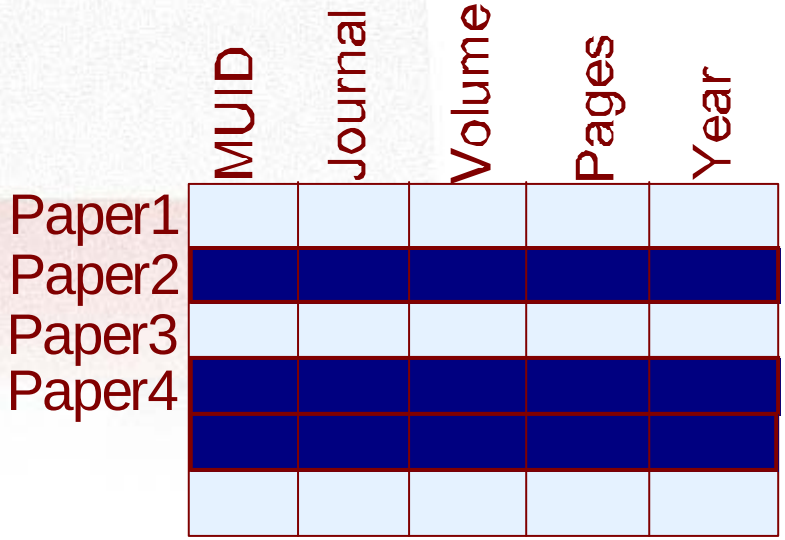
- Used to create, alter, and delete databases and tables.
- Statements
  - Create Table
    - CREATE TABLE table_name (column_name1 data_type primary key, column_name2 data_type);
  - Some other operations
    - “alter” and “drop”

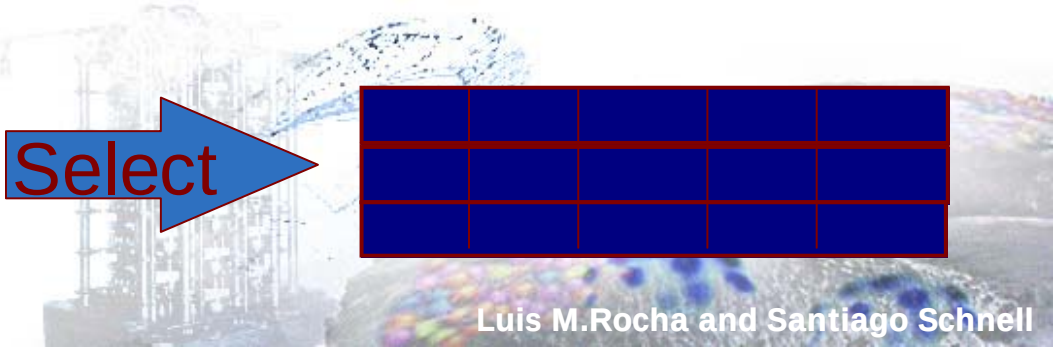
Data Manipulation Language

- Used to retrieve, insert, delete and update data in a database
- Statements
  - Select
    - Selects rows (records) according to attribute criteria
      - E.g. Select CDs published in YEAR=x
  - Some other operations
    - “insert”, “update”, “delete”, and “truncate”

Select Statement

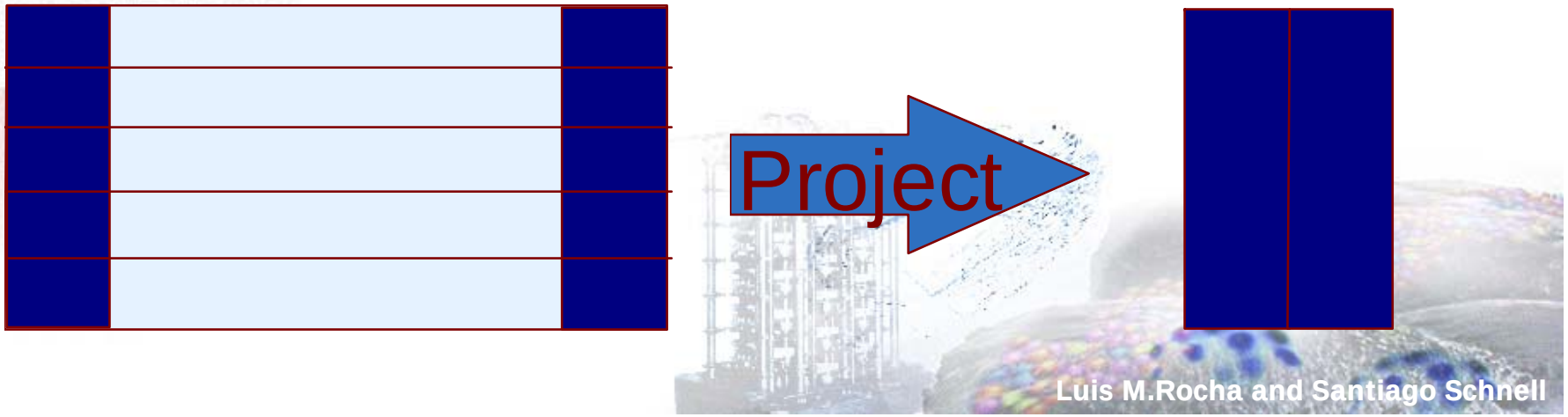
- Select
  - Selects rows (records) according to attribute criteria
    - E.g. papers published in YEAR=x
  - **SELECT** * FROM *list-of-relations* WHERE *condition*
    - **SELECT** * FROM *CITATION_TABLE* WHERE *PUBLISHED_YEAR*='1995';
  - * Denotes ALL
  - **SELECT** * FROM *T;*
    - Returns all elements of all the rows of the table T





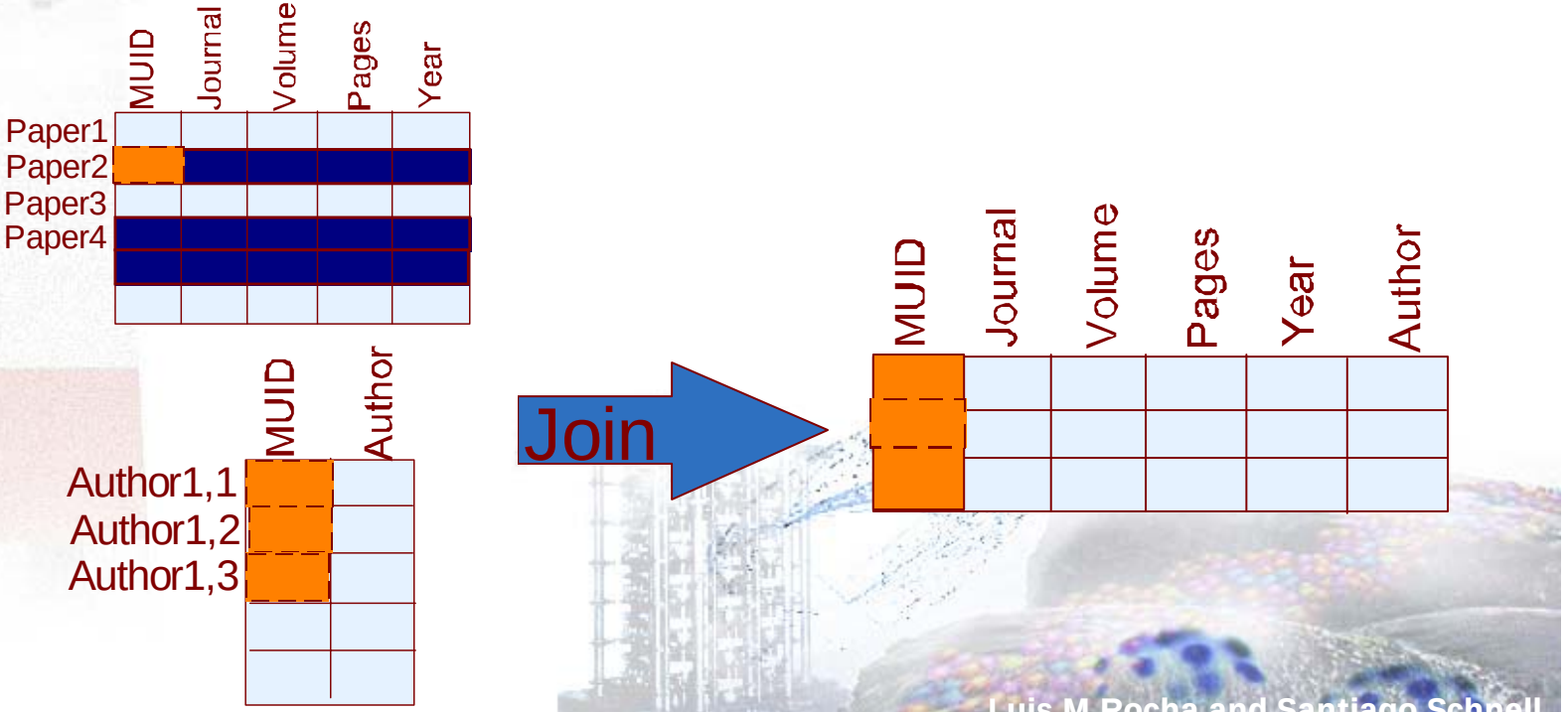
Projection Operation

- Project
  - Extracts columns
    - E.g. projects a set of papers into a reduced set of attributes.
  - **SELECT** *C1,C7* FROM *T*;



Join Operation

- Join
  - Merges records that contain matching values for specified attributes
    - given a key value join records from both tables
  - **SELECT** * FROM *employee, department*;
  - **SELECT** * FROM *citation-table, author-table* WHERE citation-table.*MUID* = author-table.*MUID*;




Betty
By Delainey & Rasmussen
...AND YOUR AUTOGRAPH, PLEASE...
8
RIGHT
MAKE IT OUT TO "CYNTHIA, MY NUMBER ONE FAN"
...AND HERE'S YOUR CUSTOMER CARD!
THANKS
HAVE A NICE DAY, MRS. SLUG
THANKS
HAVE A SAFE TRIP BACK TO 1320 WEATHERWAX CRESCENT, ENJOY A MEAL OF MACARONI AND CHEESE OR QUESADILLAS
READ A GOOD DETECTIVE NOVEL, HAVE A FRAGRANT BUBBLE BATH
WATCH "SURVIVOR" AND HIT THE SACK — BY THE WAY, YOUR HAIR COLOR GOES ON SALE NEXT WEDNESDAY
© 2001 by NEA, Inc. www.comics.com 11-18
8
7
STILL TOO MUCH?
I SAY JUST STICK WITH THEIR NAMES. USING ANY MORE INFO TENDS TO GIVE CUSTOMERS THE HEEBIE JEEBIES

Group Assignment

- Third Installment
  - Given any text such as the *library of babylon* or *Funes, the memorious*
    - Create a ***database model*** and a ***relational database instance*** using *Microsoft Access* to store the data and conclusions from previous installments
      - Use the entity-relationship model
    - Examples of items that should appear
      - Title, author, language, publication date
      - Frequency/probability of each letter
      - Conditional probabilities for letters ‘e’ and ‘u’ (as produced in installment 2)
      - Positively and negatively dependent letters
    - Use at least 4 texts
  - Due on April 27th, 2005
  - Upload to Oncourse

Next Class!

- Topics of next classes
  - Databases and SQL
  - Individual Assignment
  - Review
- Readings for Next week
  - @ *infoport*
  - course package
- No More Labs!!!!!!!