

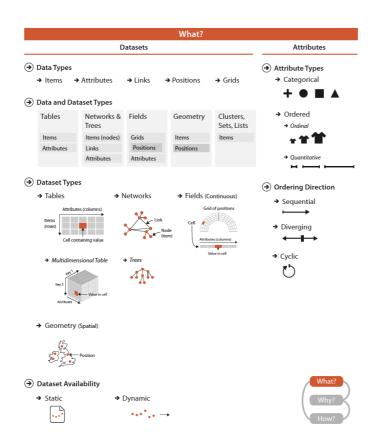
Information Visualization and Visual Analytics (M1522.000500)

What: Data Abstraction

Jinwook Seo, Ph. D.

Professor, Dept. of Computer Science and Engineering Seoul National University







Why Do Data Semantics and Types Matter?

Basil, 7, S, Pear

• What does each word mean?

→ Semantics: real-world meaning

ID	Name	Age	Shirt Size	Favorite Fruit
1	Amy	8	S	Apple
2	Basil	7	S	Pear
3	Clara	9	M	Durian
4	Desmond	13	L	Elderberry
5	Ernest	12	L	Peach
6	Fanny	10	S	Lychee
7	George	9	M	Orange
8	Hector	8	L	Loquat
9	Ida	10	M	Pear
10	Amy	12	M	Orange

Information Visualization and Visual Analytics – Data Abstractio

What: Data Abstraction



Why Do Data Semantics and Types Matter?

- Types of the data: structural or mathematical interpretation
 - Data level → Data Types
 - what kind of thing is it?
 - item, attribute, link, position, grid
 - Dataset level → Dataset Types
 - how are these data types combined into a larger structure
 - table, tree, field of sampled values Scientific visualization에서 많이 씀,
 - Attribute level → Attribute Types information vis 에서 많이 안씀
 - what kinds of math operations are meaningful for it?
 - attribute: property that can be measured, observed, or logged
 - → variable, dimension
 - Number of detergents: quantity addition/subtraction
 - Postal code: code category



Data Types

- Attribute: specific property that can be measured, observed, or logged (a.k.a variable or dimension)
- Item: individual entity that is discrete
- Link: relationship between items
- **Grid**: specifies the strategy for sampling continuous data in terms of both geometric and topological relationships between its cells
- Position: spatial data, providing a location

nformation Visualization and Visual Analytics – Data Abstractio

What: Data Abstraction



Dataset Types and Data types

- Dataset
 - collection of info that is the target of analysis
 - arise from combinations of data types

Tables	Networks & Trees	Fields	Geometry	Clusters, Sets, Lists
Items	Items (nodes)	Grids	Items	Items
Attributes	Links	Positions	Positions	
	Attributes	Attributes		



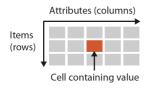
Basic Dataset Types

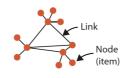
→ Tables

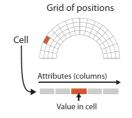
→ Networks

→ Fields (Continuous)

→ Geometry (Spatial)

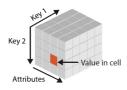








→ Multidimensional Table





→ Trees

Information Visualization and Visual Analytics – Data Abstractio

What: Data Abstraction



Dataset Type: Tables

- item / record / tuple
- attribute / field / variable / dimension
- cell contains value
 - quantitative
 - ordinal
 - nominal

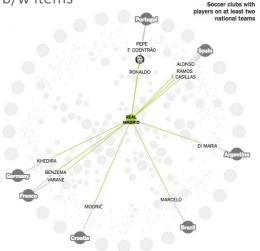
Α	В	С	S	Т	U
Order ID	Order Date	Order Priority	Product Container	Product Base Margin	Ship Date
3	10/14/06	5-Low	Large Box	0.8	10/21/06
6	2/21/08	4-Not Specified	Small Pack 0.55		2/22/08
32	7/16/07	2-High	Small Pack	0.79	7/17/07
32	7/16/07	2-High	Jumbo Box	•1	7/17/07
32	7/16/07	2-High	Medium Box	attribute	7/18/07
32	7/16/07		Medium Box	0.03	7/18/07
35	10/23/07	4-Not Specified	Wrap Bag	0.52	10/24/07
35	10/23/07	4-Not Specified	Small Box	0.58	10/25/07
36	11/3/07	1-Urgent	Small Box	0.55	11/3/07
65		1-Urgent	Small Pack	0.49	3/19/07
66	1 /20 /05	5-Low	Wrap Bag	0.56	1/20/05
69		4-Not Specified	Small Pack	0.44	6/6/05
69	1001115	4-Not Specified	Wrap Bag	0.6	6/6/05
70	12/18/06	5-Low	Small Box	0.59	12/23/06
70	12/18/06	5-Low	Wrap Bag	0.82	12/23/06
96	4/17/05	2-High	Small Box	0.55	4/19/05
97	1/29/06	3-Medium	Small Box	0.38	1/30/06
129	11/19/08	5-Low	Small Box 0.37		11/28/08
130	5/8/08	2-High	Small Box 0.37		5/9/08
130	5/8/08	2-High	Medium Box	0.38	5/10/08
130	5/8/08	2-High	Small Box	0.6	5/11/08
132	6/11/06	3-Medium	Medium Box	0.6	6/12/06
132	6/11/06	3-Medium	Jumbo Box	0.69	6/14/06
134	5/1/08	4-Not Specified	Large Box	0.82	5/3/08
135	10/21/07	4-Not Specified	Small Pack	0.64	10/23/07
166	9/12/07	2-High	Small Box	0.55	9/14/07
193		1-Urgent	Medium Box	0.57	8/10/06
194	4/5/08	3-Medium	Wrap Bag	0.42	4/7/08

• multidimensional table: indexed with multiple keys



Dataset Type: Network and Trees

- Well suited when there is some kind of relationship b/w items
- Node: item
 - Can have associated attributes
- Link: relation between two items
- Trees: networks with a hierarchical structure
 - Unlike network, there are no cycles



Information Visualization and Visual Analytics – Data Abstraction

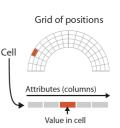
What: Data Abstraction

hci lab

Dataset Type: Fields

- Contains data created by sampling on Grids
- Cell contains measurements from continuous domain
 - sampling, interpolation & reconstruction
- Spatial Fields: cell structure of the field is based on sampling at spatial positions
 - Nonspatial data: abstract data
 - Scivis: spatial position is **given** with the dataset
 - · Handling continuous data appropriately within the math framework of signal processing
 - Infovis: the use of space in visual encoding is **chosen** by the designer
 - The chosen idiom is suitable for the combination of data and task

Fields (Continuous)

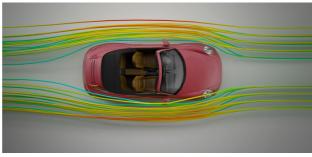




Dataset Type: Fields - Spatial Fields

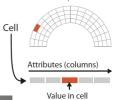
- Spatial Fields: cell structure of the field is based on sampling at spatial positions
 - Scivis: spatial position is given with the dataset
 - Handling continuous data appropriately within the math framework of signal processing





Fields (Continuous)

Grid of positions



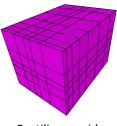
nformation Visualization and Visual Analytics – Data Abstractio

What: Data Abstraction

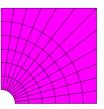


Dataset Type: Fields - Grid Types

- Grid needs
 - geometry: location in space
 - topology: how each cell connects with its neighboring cells
- Uniform grid: sampling at completely regular intervals
 - No need of grid geometry and topology
- Rectilinear grid: supports nonlinear sampling
 - some info about geometric location of each row
- Structured grid: allows curvilinear shapes
 - geometric location of each cell should be specified
- Unstructured grid: allows complete flexibility
 - topological & geometric information required

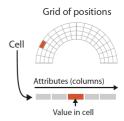


Rectilinear grid



Curvilinear grid

Fields (Continuous)





Dataset Type: Geometry

- Specifies info about the shape of items with explicit spatial positions
- Often includes hierarchical structure at multiple scales
 - May be intrinsic or may be derived from the original data
- Not necessarily have attributes
- Simply showing a geometric dataset is not an interesting problem for a vis designer

Geometry (Spatial)



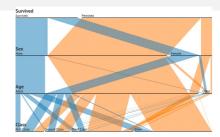
Information Visualization and Visual Analytics – Data Abstraction

What: Data Abstraction



Dataset Type: Combinations

- Set: unordered group of items
- List: ordered group of items
- Cluster: grouping based on attribute similarity
- Path: ordered set + links connecting nodes
- Compound network: network with an associated tree



		Para and a second	-	
to Na	11 my		WER	SNE
test a Satural Science Chair a Dig 1 Stort	Property Land Land States	DAY I HAVE BEEN	Chief Said	the about the state of
	1 7 7 T			
" priese pape page 1816 jahr.	and the same of the	Page 1 and a second	done one in 7	OLA TABLESHIE METAL
Mr 2 170		The Root of Lane		in Againmington
(top card) have the	一直 通月 カーオー	74 1/1		1.5
NAME AND ADDRESS OF THE PARTY O	Per Per Killer	Fifth Townson	WHEN LOOK HALL	Washing Library
MAX TO SERVICE STATE OF THE SERVICE STATE STA	LEVER MANAGEMENT	THE LANS.	on B (887 10	THE PERSON
to I you had pact hot leave	IN REPORT OF THE			540 PO
	THE STREET	ALMOS DE LOS	Office Services	berter displaced to
NAME OF TAXABLE PARTY.	Transference and the later	Married Street, Square,	excession of the	TOP .
covering a browning larger secon	I de la	or transfer but		on the world der trail, Property
	Indiana in factor	A ALCOHOL:	* × ×	E-
AND MALE OF SAME AND AND	THE RESERVE	A 700 mg	Oper Notice	and plant for less;
THE PARTY NAMED IN	CHAPTER SEC	V 71.71	11/18	BE PR THE
Section of the section of the	In this entire state of	WINDS OF	physic har he	and the party of the last
Water Street Street in	一十十十十二十二十二十二十二十二十二十二十二十二十二十二十二十二十二十二十二十	A Personal Property lies	Street, respect	the way with all property and
- BI (BL (810'))	1-11-11-11-11	The second		
A part particular of the	MARKS THE STREET	1000	April of the Paris	TANKS THE STREET NAME AND THE PARTY
9800	HARRY MARKETON		A Part of the last	
her Kirken IIII haddale levil	44 C T T T T T T T T T T T T T T T T T T	100000	William Inc.	NAMES OF THE PARTY OF
	And the second	The second	And I	THE PERSON
BIR OF STREET	THE PROPERTY.	CD20X	Japan 100	The latest
National States	THE RESERVOIR	No. of the Co.	W. N.	THE COLUMN
The last water to	Contract of the last	A SHOW	SCHOOL STORY	Committee of the Party of the Party of
Sale Sanding Bell	- 1 15 D B S S S S S S	Section 1	200	
	4.1.044	The same of		200
- mRu	THE PERSON NAMED IN COLUMN		-	plant of the best of the last
THE R. LEWIS CO., LANSING MICH. P. LEWIS CO., LANSING, MICH. P. LEWIS CO.,	- T. LOWER BOOK	10 Sec. 1991	THE REAL PROPERTY.	1 1 1 1 1 1 1 1 1
and the second	-1 200 00 C	400000	and the same of	The same of the sa
Sale has always	- March 1997		and the contract of	4 / 11
The same of the	The second second second	1	Section 1	PROPERTY.
194 × 1005	Section 1	NAME OF TAXABLE PARTY.	ALC: NAME OF TAXABLE PARTY.	heptions / //
V bykobyky	- Land	Sept Tribut		Stat 901 No.
-Decomp	out of the Street	The Physics		Delica (Su

Figure 1: Directory structure of a Web site visualized as a Treemap with external links overlaid as curves. Blue curves are HTML links, red curves are image links.



Abstraction and Availability

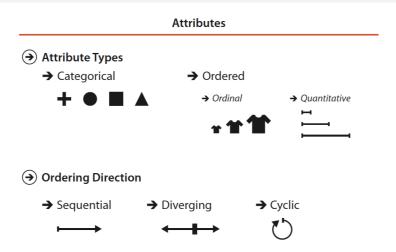
- Data Abstraction
 - domain-specific to GENERIC
 - translate domain-specific terms into words that are as generic as possible
- Data Availability
 - Static File: available all at once
 - Dynamic Streams

nformation Visualization and Visual Analytics – Data Abstractio

What: Data Abstraction



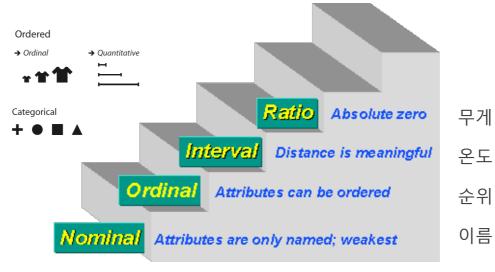
Attribute Types



• Hierarchical Attributes: within an attribute or between multiple attributes



Levels of Measurements



http://www.socialresearchmethods.net/kb/measlevl.php

Information Visualization and Visual Analytics – Data Abstraction

What: Data Abstraction



Semantics - Key vs. Value Semantics

- Key attribute acts as an index used to look up value attributes
- Flat Tables: only one key
 - Implicit key: keys are simply the index of the row

A	В	C	S	T	U
Order ID	Order Date	Order Priority	Product Container	Product Base Margin	Ship Date
3	10/14/06	5-Low	Large Box	0.8	10/21/06
6	2/21/08	4-Not Specified	Small Pack	0.55	2/22/08
32	7/16/07	2-High	Small Pack	0.79	7/17/07
32	7/16/07	2-High	Jumbo Box	0.72	7/17/07
32	7/16/07	2-High	Medium Box	0.6	7/18/07
32	7/16/07	2-High	Medium Box	0.65	7/18/07
35	10/23/07	4-Not Specified	Wrap Bag	0.52	10/24/07
35	10/23/07	4-Not Specified	Small Box	0.58	10/25/07
36	11/3/07	1-Urgent	Small Box	0.55	11/3/07
65	3/18/07	1-Urgent	Small Pack	0.49	3/19/07
00	1 /20 /05	F.L.	Miles Dee	0.50	1 /00 /05

구분을 안함 multidimensional/flat

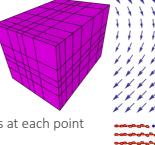
- Explicit key: keys may be categorical or ordinal attributes (unique)
- Multidimensional Tables: multiple keys required
 - combination of all keys must be unique for each item, even though an individual key may contain duplicates
 - independent keys vs. dependent values

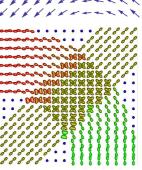


Semantics - Key vs. Value Semantics - Fields

의미가 조금 다름

- Fields differ from tables (Fields are characterized in terms of # of keys vs. values)
 - Each cell represents **continuous** data
 - Key, value → independent, dependent variable (spatial field data)
 - Multivariate structure depends on # of value attributes
 - Multidimensional structure depends on # of key attributes
- Scalar Field: univariate one attribute per cell (e.g., CT scan)
- Vector Fields: multivariate with a list of multiple attribute values at each point
- Tensor Fields:
 - Representing more complex multivariate mathematical structure with an array of attributes at each point
 - e.g., stress in 3D field: defined by nine numbers that represent forces acting in three orthogonal directions → can be represented as a more complex shape, e.g., ellipsoid.





Information Visualization and Visual Analytics – Data Abstraction

What: Data Abstraction

hci ldb seoul national university

Semantics - Temporal Semantics

- Temporal attribute: any info relates to time
- Complicated to handle: rich hierarchical structure
 - Multiscale: nanoseconds ~ decades
 - Weeks do not cleanly fit into months
- Time-varying semantic: time is one of the key attributes
 - Time-series dataset: an ordered sequence of time-value pairs
 - Not always spaced at uniform intervals
- Dynamic
 - Time-varying semantics: time is a key attribute
 - Stream type dataset: items can change during a running session



Note

• Questions?

Information Visualization and Visual Analytics – Data Abstraction