Unit 4: Databases

Dave Abel

February 26th, 2016

Table Format

Product ID	Company	Prime	Price	# Ratings	Avg. Rating
4615253	Winsome Wood	TRUE	26.97	1,279	4
3249814	Crown Mark	TRUE	37.41	1,675	4
1235561	Lifetime	TRUE	51.88	931	4.5
4224	Lifetime	TRUE	31.58	1,003	5

Table: Primary Key

Product ID	Company	Prime	Price	# Ratings	Avg. Rating
4615253	Winsome Wood	TRUE	26.97	1,279	4
3249814	Crown Mark	TRUE	37.41	1,675	4
1235561	Lifetime	TRUE	51.88	931	4.5
4224	Lifetime	TRUE	31.58	1,003	5

Database Queries: *SQL*

- Examples:

```
SELECT company FROM alabamazon;
```

Product ID	Company	Prime	Price	# Ratings	Avg. Rating
4615253	Winsome Wood	TRUE	26.97	1,279	4
3249814	Crown Mark	TRUE	37.41	1,675	4
1235561	Lifetime	TRUE	51.88	931	4.5
4224	Lifetime	TRUE	31.58	1,003	5

Database Queries: *SQL*

- Examples:

```
SELECT company FROM alabamazon;
```

Company
Winsome Wood
Crown Mark
Lifetime
Lifetime

Database Queries: *SQL*

- Examples:

```
SELECT company FROM alabamazon where rating=“4”;
```

Product ID	Company	Prime	Price	# Ratings	Avg. Rating
4615253	Winsome Wood	TRUE	26.97	1,279	4
3249814	Crown Mark	TRUE	37.41	1,675	4
1235561	Lifetime	TRUE	51.88	931	4.5
4224	Lifetime	TRUE	31.58	1,003	5

Database Queries: *SQL*

- Examples:

```
SELECT company FROM alabamazon where rating=“4”;
```

Product ID	Company	Prime	Price	# Ratings	Avg. Rating
4615253	**Winsome Wood**	**TRUE**	**26.97**	**1,279**	**4**
3249814	**Crown Mark**	**TRUE**	**37.41**	**1,675**	**4**
1235561	Lifetime	TRUE	51.88	931	4.5
4224	Lifetime	TRUE	31.58	1,003	5

Database Queries: *SQL*

- Examples:

```
SELECT company FROM alabamazon;
```

Company
Winsome Wood
Crown Mark

Database Queries: *SQL*

- Examples: *aggregating functions*

```
SELECT count(*) FROM alabamazon where rating > 4;
```

Product ID	Company	Prime	Price	# Ratings	Avg. Rating
4615253	Winsome Wood	TRUE	26.97	1,279	4
3249814	Crown Mark	TRUE	37.41	1,675	4
1235561	Lifetime	TRUE	51.88	931	4.5
4224	Lifetime	TRUE	31.58	1,003	5

Database Queries: *SQL*

- Examples: *aggregating functions*

```
SELECT count(*) FROM alabamazon where rating > 4;
```

Reports: 2

Product ID	Company	Prime	Price	# Ratings	Avg. Rating
4615253	Winsome Wood	TRUE	26.97	1,279	4
3249814	Crown Mark	TRUE	37.41	1,675	4
1235561	**Lifetime**	**TRUE**	**51.88**	**931**	**4.5**
4224	**Lifetime**	**TRUE**	**31.58**	**1,003**	**5**

Replicated Database

- Idea: store the database more than once in other secure locations!

Product ID	Company	Prime	Price	# Ratings	Avg. Rating
4615253	Winsome Wood	TRUE	26.97	1,279	4
3249814	Crown Mark	TRUE	37.41	1,675	4
1235561	Lifetime	TRUE	51.88	931	4.5
4224	Lifetime	TRUE	31.58	1,003	5

Replicated Database

- Idea: store the database more than once in other secure locations!

Product ID	Company	Prime	Price	# Ratings	Avg. Rating
4615253	Winsome Wood	TRUE	26.97	1,279	4
3249814	Crown Mark	TRUE	37.41	1,675	4
1235561	Lifetime	TRUE	51.88	931	4.5
4224	Lifetime	TRUE	31.58	1,003	5

Replicated Database

- Idea: store the database more than once in other secure locations!

Product ID	Company	Prime	Price	# Ratings	Avg. Rating
4615253	Winsome Wood	TRUE	26.97	1,279	4
3249814	Crown Mark	TRUE	37.41	1,675	4
1235561	Lifetime	TRUE	51.88	931	4.5
4224	Lifetime	TRUE	31.58	1,003	5

Replicated Database

- Idea: store the database more than once in other secure locations!

Product ID	Company	Prime	Price	# Ratings	Avg. Rating
4615253	Winsome Wood	TRUE	26.97	1,279	4
3249814	Crown Mark	TRUE	37.41	1,675	4
1235561	Lifetime	TRUE	51.88	931	4.5
4224	Lifetime	TRUE	31.58	1,003	5

Replicated Database

- One Idea: only change both.

Add: “82374, Dave’s Chairs, FALSE, 450, 1, 5”

Product ID	Company	Prime	Price	# Ratings	Avg. Rating
4615253	Winsome Wood	TRUE	26.97	1,279	4
3249814	Crown Mark	TRUE	37.41	1,675	4
1235561	Lifetime	TRUE	51.88	931	4.5
4224	Lifetime	TRUE	31.58	1,003	5

In Providence, RI

Product ID	Company	Prime	Price	# Ratings	Avg. Rating
4615253	Winsome Wood	TRUE	26.97	1,279	4
3249814	Crown Mark	TRUE	37.41	1,675	4
1235561	Lifetime	TRUE	51.88	931	4.5
4224	Lifetime	TRUE	31.58	1,003	5

In Portland, OR

Replicated Database

- One Idea: only change both.

Add: “82374, Dave’s Chairs, FALSE, 450, 1, 5”

Product ID	Company	Prime	Price	# Ratings	Avg. Rating
4615253	Winsome Wood	TRUE	26.97	1,279	4
3249814	Crown Mark	TRUE	37.41	1,675	4
1235561	Lifetime	TRUE	51.88	931	4.5
4224	Lifetime	TRUE	31.58	1,003	5
82734	Dave’s Chairs	FALSE	450	1	5

In Providence, RI

Product ID	Company	Prime	Price	# Ratings	Avg. Rating
4615253	Winsome Wood	TRUE	26.97	1,279	4
3249814	Crown Mark	TRUE	37.41	1,675	4
1235561	Lifetime	TRUE	51.88	931	4.5
4224	Lifetime	TRUE	31.58	1,003	5
82734	Dave’s Chairs	FALSE	450	1	5

In Portland, OR

Two Way Database

- Databases can often carry *redundant* information.
- A Two Way database is a way to consolidate.

Two Way Database

Product ID	Company	Company Ships to Canada	Price	# Ratings	Avg. Rating	Company Verified
4615253	Winsome Wood	TRUE	26.97	1,279	4	TRUE
3249814	Crown Mark	TRUE	37.41	1,675	4	TRUE
1235561	Lifetime	TRUE	51.88	931	4.5	TRUE
4224	Lifetime	TRUE	31.58	1,003	5	TRUE
129384	Dave’s Chairs	FALSE	450	1	5	FALSE
33332341	Dave’s Chairs	FALSE	27	948	5	FALSE
9723	Dave’s Chairs	FALSE	12	44	5	FALSE
761	Dave’s Chairs	FALSE	900	2	5	FALSE

Two Way Database

Product ID	Company	Company Ships to Canada	Price	# Ratings	Avg. Rating	Company Verified
4615253	Winsome Wood	TRUE	26.97	1,279	4	TRUE
3249814	Crown Mark	TRUE	37.41	1,675	4	TRUE
1235561	**Lifetime**	**TRUE**	51.88	931	4.5	**TRUE**
4224	**Lifetime**	**TRUE**	31.58	1,003	5	**TRUE**
129384	**Dave's Chairs**	**FALSE**	450	1	5	**FALSE**
33332341	**Dave's Chairs**	**FALSE**	27	948	5	**FALSE**
9723	**Dave's Chairs**	**FALSE**	12	44	5	**FALSE**
761	**Dave's Chairs**	**FALSE**	900	2	5	**FALSE**

Two Way Database

Product ID	Company	Price	# Ratings	Avg. Rating
4615253	Winsome Wood	26.97	1,279	4
3249814	Crown Mark	37.41	1,675	4
1235561	Lifetime	51.88	931	4.5
4224	Lifetime	31.58	1,003	5
129384	Dave's Chairs	450	1	5
33332341	Dave's Chairs	27	948	5
9723	Dave's Chairs	12	44	5
761	Dave's Chairs	900	2	5

Company	Ships to Canada	Verified
Winsome Wood	TRUE	TRUE
Crown Mark	TRUE	TRUE
Lifetime	TRUE	TRUE
Dave's Chairs	FALSE	FALSE

Two Way Database

Product ID	Company	Price	# Ratings	Avg. Rating
4615253	Winsome Wood	26.97	1,279	4
3249814	Crown Mark	37.41	1,675	4
1235561	Lifetime	51.88	931	4.5
4224	Lifetime	31.58	1,003	5
129384	Dave's Chairs	450	1	5
33332341	Dave's Chairs	27	948	5
9723	Dave's Chairs	12	44	5
761	Dave's Chairs	900	2	5

Company	Ships to Canada	Verified
Winsome Wood	TRUE	TRUE
Crown Mark	TRUE	TRUE
Lifetime	TRUE	TRUE
Dave's Chairs	FALSE	FALSE

Same info, less storage!

Clicker Question!

Episode #	Year	Host	Category	Answer
1	1984	Alex Trebek	CEREAL	Frosted Flakes
1	1984	Alex Trebek	PINK STARFISH	Patrick
1	1984	Alex Trebek	CEREAL	Raisin Brain
2	2005	Petunia Dursley	KINDS OF WOOD	Oak
2	2005	Petunia Dursley	KINDS OF WOOD	Birch
2	2005	Petunia Dursley	PLACES IN RHODE ISLAND	Providence
2	2005	Petunia Dursley	PLACES IN RHODE ISLAND	Pawtucket
3	2016	Dave Abel	DAVE'S OFFICE NUMBER	321

Q: The above table has 40 cells. How many cells *combined* would our smallest two way table have?

Clicker Question!

5 columns

8 rows

Episode #	Year	Host	Category	Answer
1	1984	Alex Trebek	CEREAL	Frosted Flakes
1	1984	Alex Trebek	PINK STARFISH	Patrick
1	1984	Alex Trebek	CEREAL	Raisin Brain
2	2005	Petunia Dursley	KINDS OF WOOD	Oak
2	2005	Petunia Dursley	KINDS OF WOOD	Birch
2	2005	Petunia Dursley	PLACES IN RHODE ISLAND	Providence
2	2005	Petunia Dursley	PLACES IN RHODE ISLAND	Pawtucket
3	2016	Dave Abel	DAVE'S OFFICE NUMBER	321

Q: The above table has 40 cells. How many cells *combined* would our smallest two way table have?

Clicker Question!

5 columns

Episode #	Year	Host	Category	Answer
1	1984	Alex Trebek	CEREAL	Frosted Flakes
1	1984	Alex Trebek	PINK STARFISH	Patrick
1	1984	Alex Trebek	CEREAL	Raisin Brain
2	2005	Petunia Dursley	KINDS OF WOOD	Oak
2	2005	Petunia Dursley	KINDS OF WOOD	Birch
2	2005	Petunia Dursley	PLACES IN RHODE ISLAND	Providence
2	2005	Petunia Dursley	PLACES IN RHODE ISLAND	Pawtucket
3	2016	Dave Abel	DAVE'S OFFICE NUMBER	321

8 rows

Q: The above table has 40 cells. How many cells *combined* would our smallest two way table have?

Hint: which column(s) are redundant?

Clicker Question!

5 columns

8 rows

Episode #	Year	Host	Category	Answer
1	1984	Alex Trebek	CEREAL	Frosted Flakes
1	1984	Alex Trebek	PINK STARFISH	Patrick
1	1984	Alex Trebek	CEREAL	Raisin Brain
2	2005	Petunia Dursley	KINDS OF WOOD	Oak
2	2005	Petunia Dursley	KINDS OF WOOD	Birch
2	2005	Petunia Dursley	PLACES IN RHODE ISLAND	Providence
2	2005	Petunia Dursley	PLACES IN RHODE ISLAND	Pawtucket
3	2016	Dave Abel	DAVE'S OFFICE NUMBER	321

[A] 21

[B] 22

[C] 26

[D] 32

Q: The above table has 40 cells. How many cells *combined* would our smallest two way table have?

Hint: which column(s) are redundant?

Clicker Answer!

Episode #	Year	Host
1	1984	Alex Trebek
2	2006	Petunia Dursley
3	2016	Dave Abel

Episode #	Category	Answer
1	CEREAL	Frosted Flakes
1	PINK STARFISH	Patrick
1	CEREAL	Raisin Brain
2	KINDS OF WOOD	Oak
2	KINDS OF WOOD	Birch
2	PLACES IN RHODE ISLAND	Providence
2	PLACES IN RHODE ISLAND	Pawtucket
3	DAVE'S OFFICE NUMBER	321

[A] 21

[B] 22

[C] 26

[D] 32

Dictionaries

Word	Definition
Cottage	A small house, especially in a countryside
Hummus	A delicious sauce made from Garbanzo Beans
Brackish	(of water) A mixture of river water and sea water

Dictionaries

Word	Definition
Cottage	A small house, especially in a countryside
Hummus	A delicious sauce made from Garbanzo Beans
Brackish	(of water) A mixture of river water and sea water

Definition: a *dictionary* is a database with one primary key, and only one other column.

Dictionaries

Word	Definition
Cottage	A small house, especially in a countryside
Hummus	A delicious sauce made from Garbanzo Beans
Brackish	(of water) A mixture of river water and sea water

Useful any time we want to associate one piece of information with one other piece of information

Dictionaries

KEY (sound)	***VALUE***
“Siri, what is the weather like?”	<Query the weather>
“Siri, what is the date?”	<Query the date>
“Siri, what is your name?”	<Query response about my name>

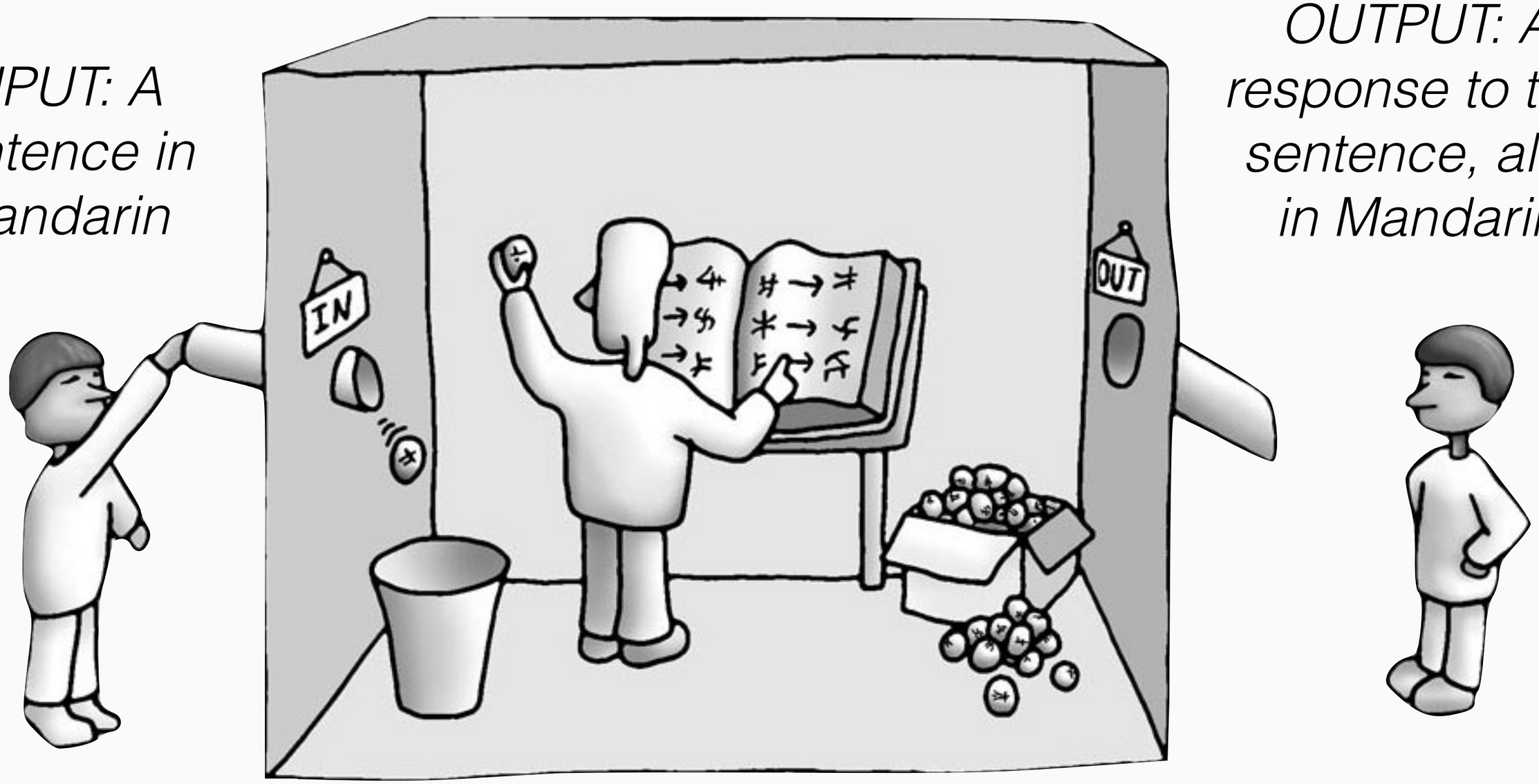
With a big enough table…

We can basically do this for any data type, as we saw in unit 1!

So we can get a “dictionary” where any data type is the key!

The Chinese Room

INPUT: A sentence in Mandarin



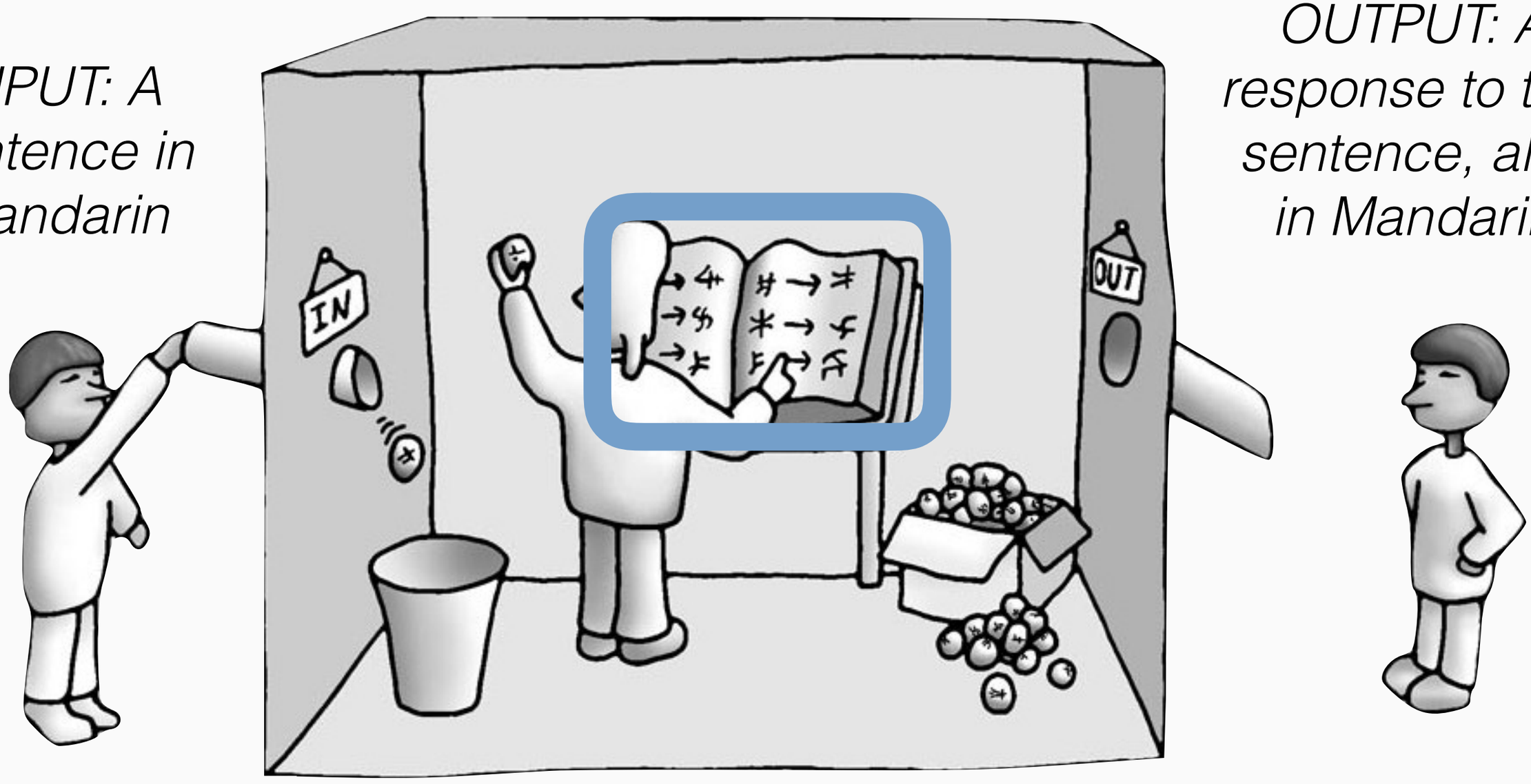
OUTPUT: A response to that sentence, also in Mandarin

http://america.pink/images/9/6/3/2/5/4/en/2-chinese-room.jpg

[Searle 1980]

The Chinese Room

INPUT: A sentence in Mandarin



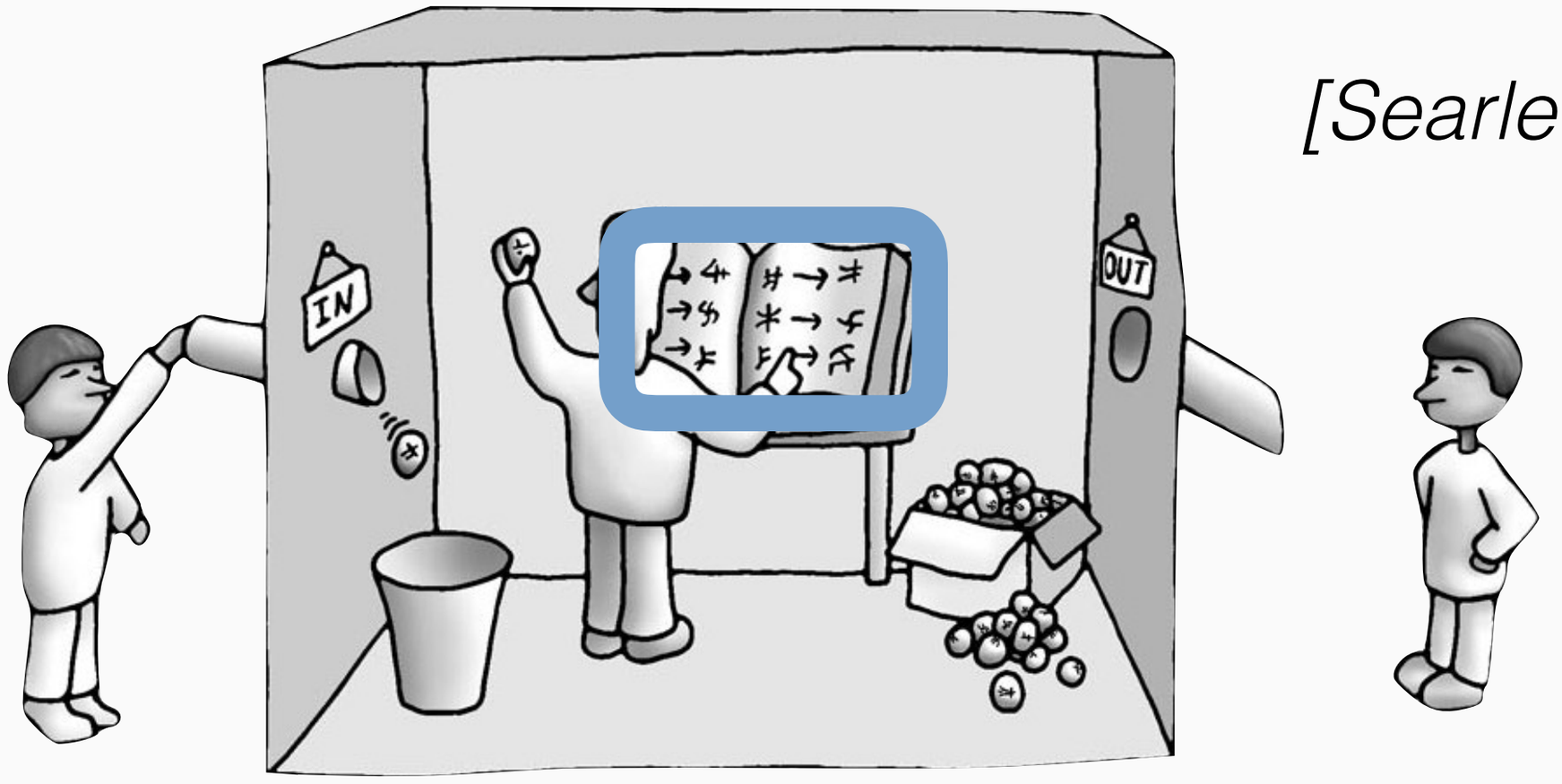
OUTPUT: A response to that sentence, also in Mandarin

http://america.pink/images/9/6/3/2/5/4/en/2-chinese-room.jpg

[Searle 1980]

The Chinese Room

[Searle 1980]

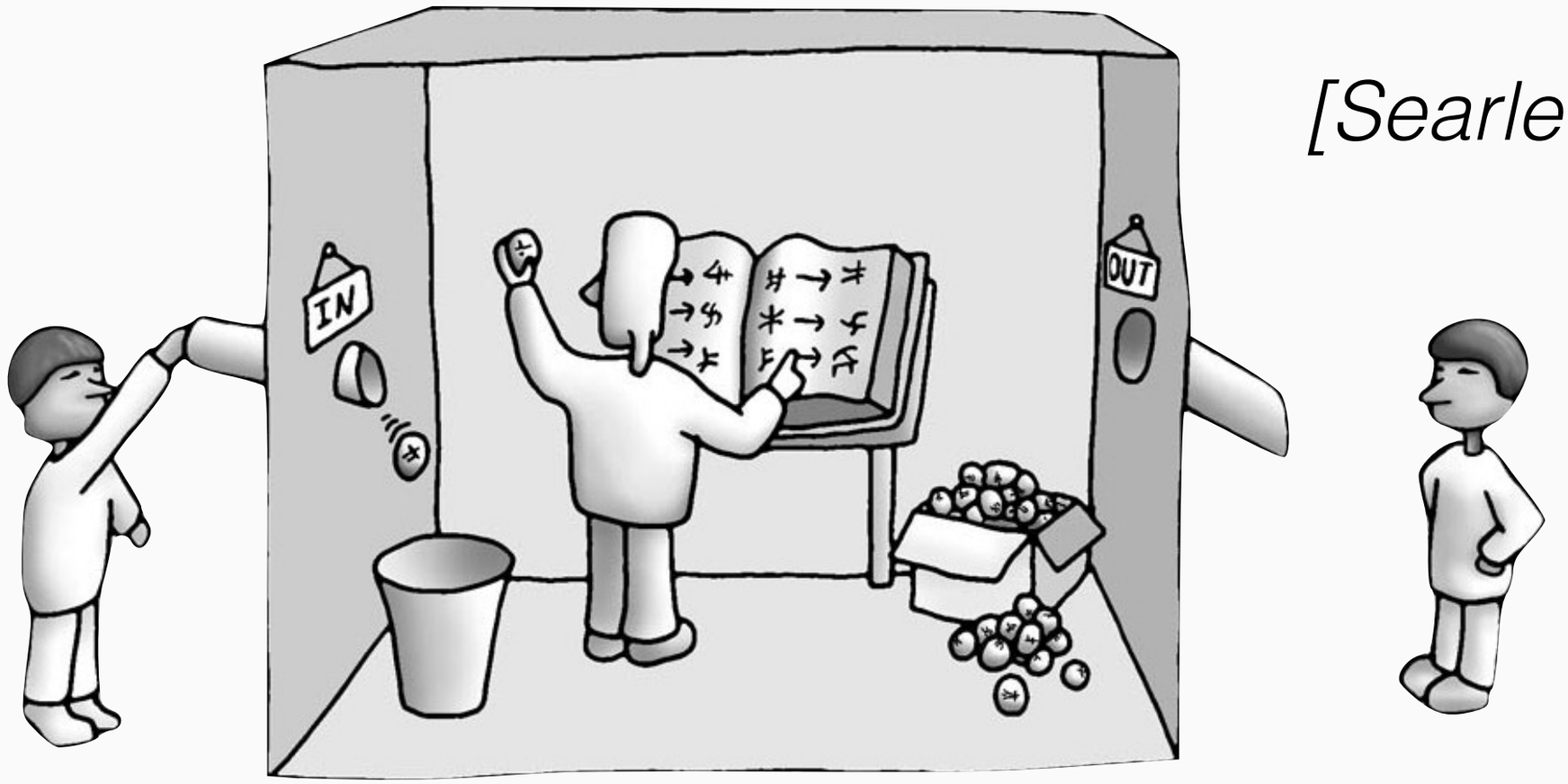


http://america.pink/images/9/6/3/2/5/4/en/2-chinese-room.jpg

Argument: Computers can never “understand” language?

Clicker Question!

[Searle 1980]



http://america.pink/images/9/6/3/2/5/4/en/2-chinese-room.jpg

[A] Computers will be able to understand language

[B] Computers will NOT be able to understand language

Dictionaries: Scratch!

Let's take a look!

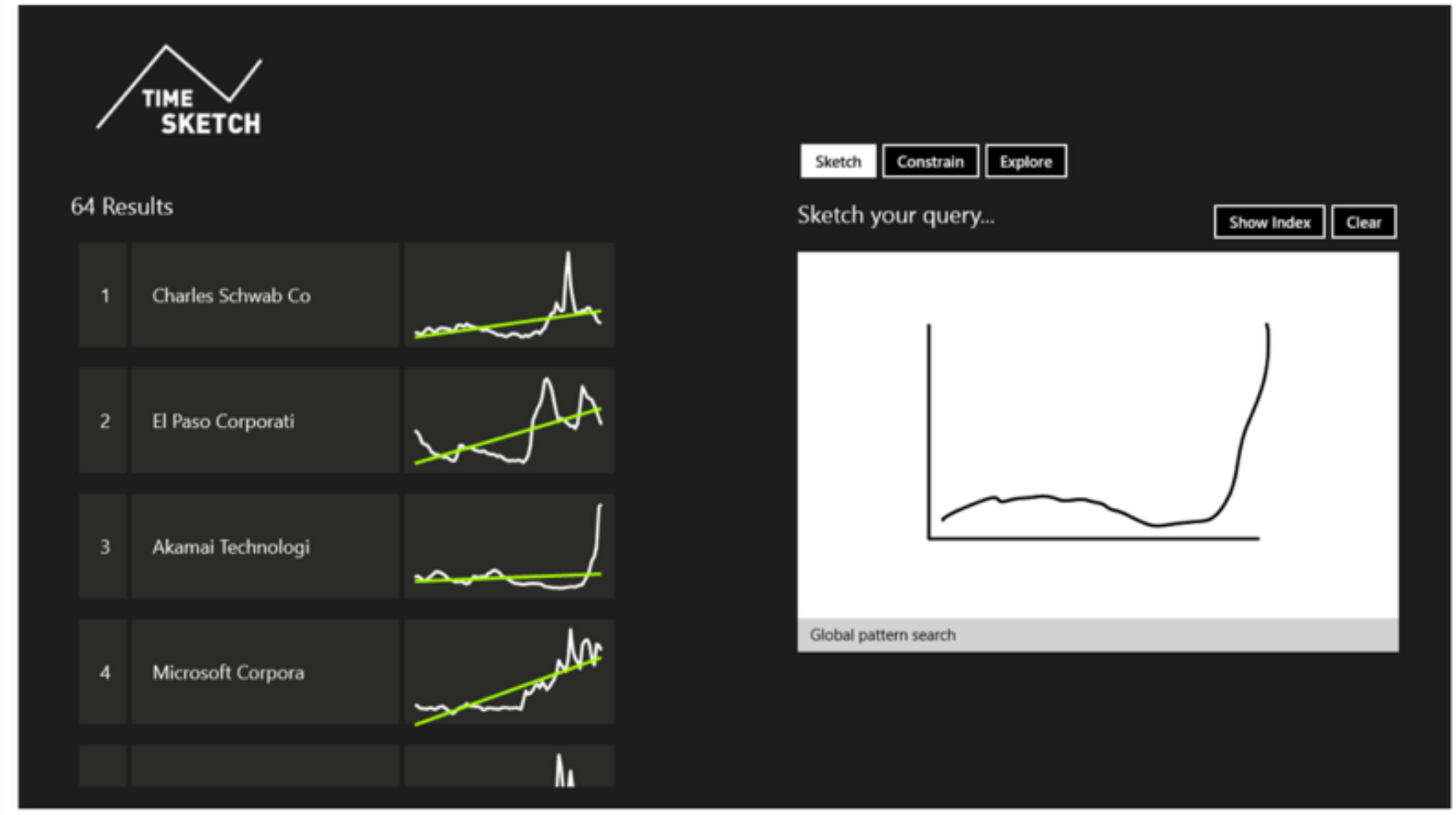
Reflection

- Databases are basically tables!
- Tables have *primary keys*
- Transactions must be used carefully..
- SQL
- Replicated Databases
- Two Way Database
- Dictionary

Visualizations!

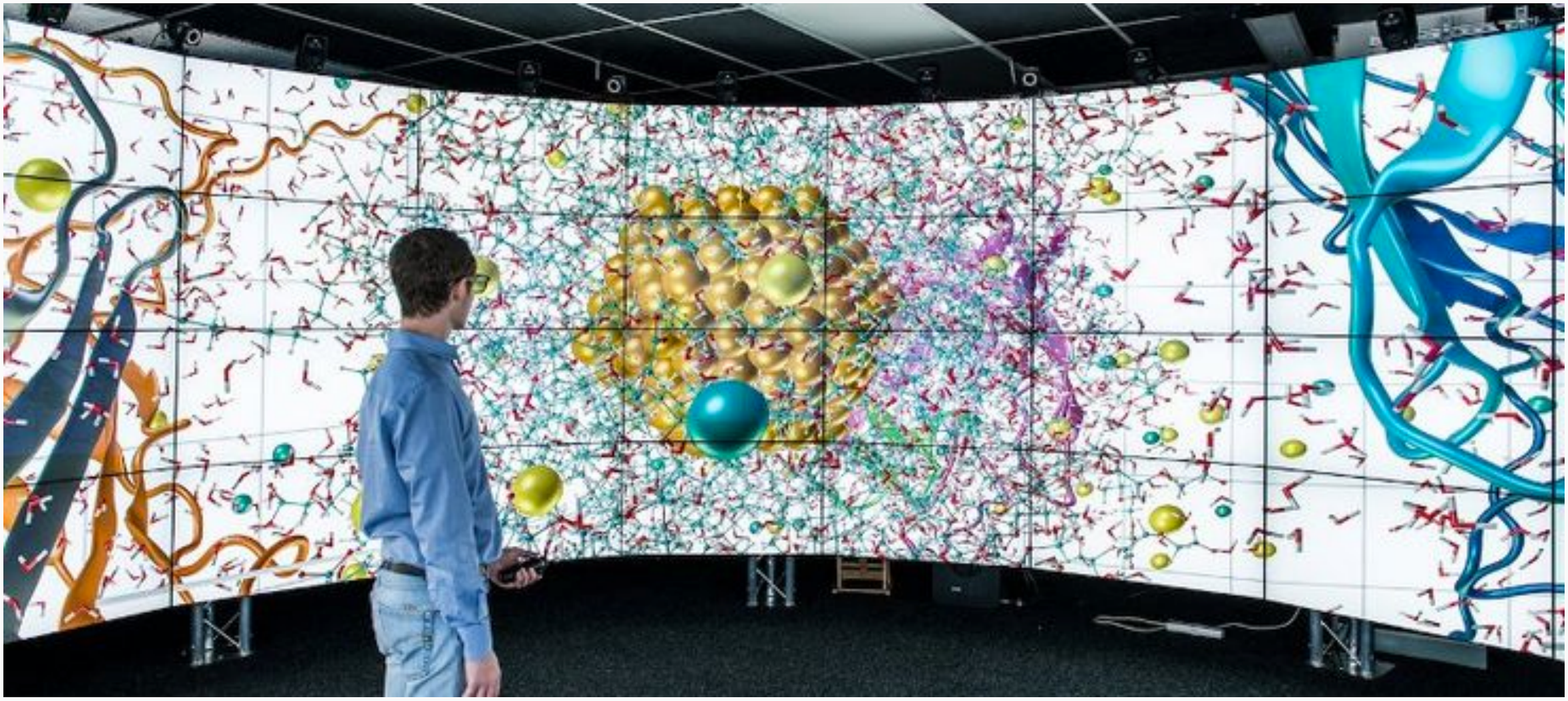
- Computers can read these big tables effectively…
- Not as easy for humans.
- Solution: Data Visualization!

Interact Visually!



[Eichmann '15]

Visualizing Data: Virtual Reality



Visualizing Data

- Current wind on Earth
- NYC Sunsets
- Biological Neural Network
- NBA shot chart visualizer
- NBA shot charts, Python
- Fashion Week

Datasets for Research!

- In Artificial Intelligence, datasets have led to *massive* advancements, especially in Computer Vision and Natural Language Processing.

- WordNet, BabelNet
- ImageNet, SUN, MSCOCO
- Average Images

Crowdsourcing

Crowdsourcing

Flower!

Plant!

Garden!

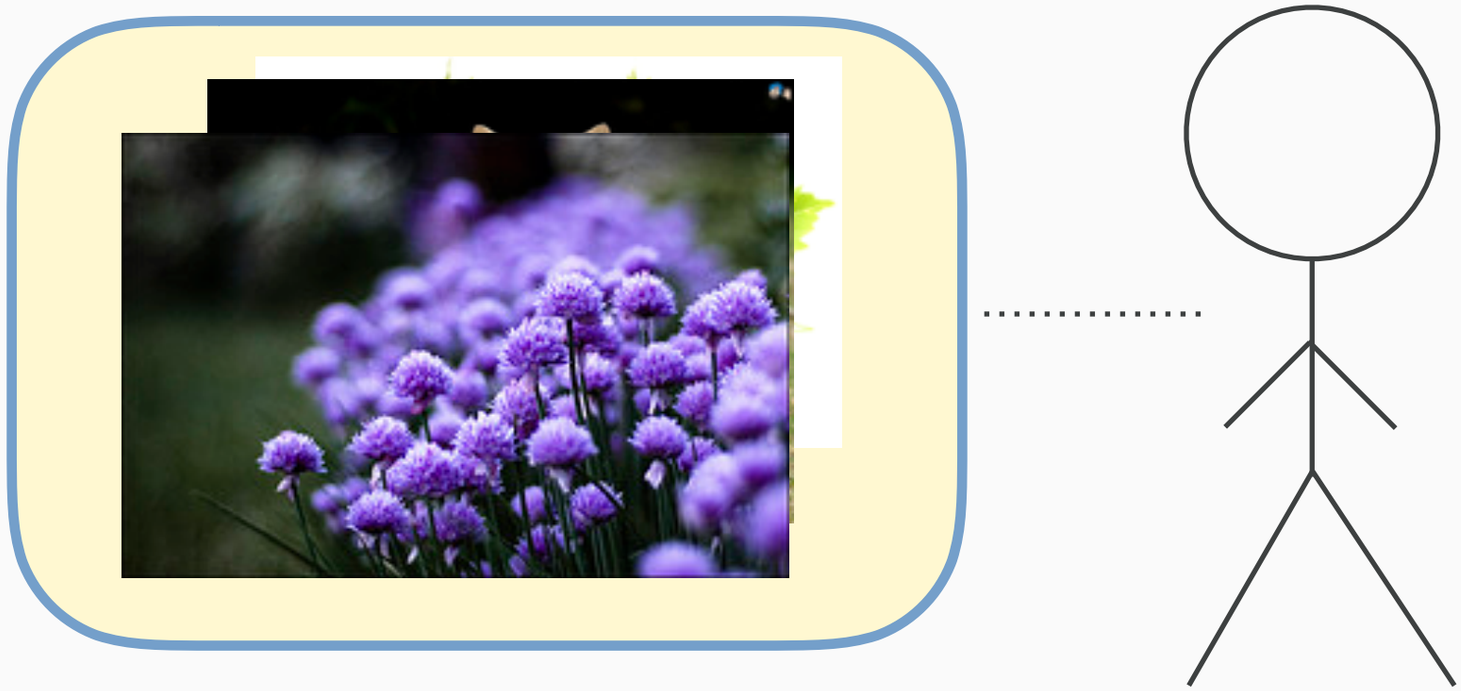
Crowdsourcing

Flower!

Plant!

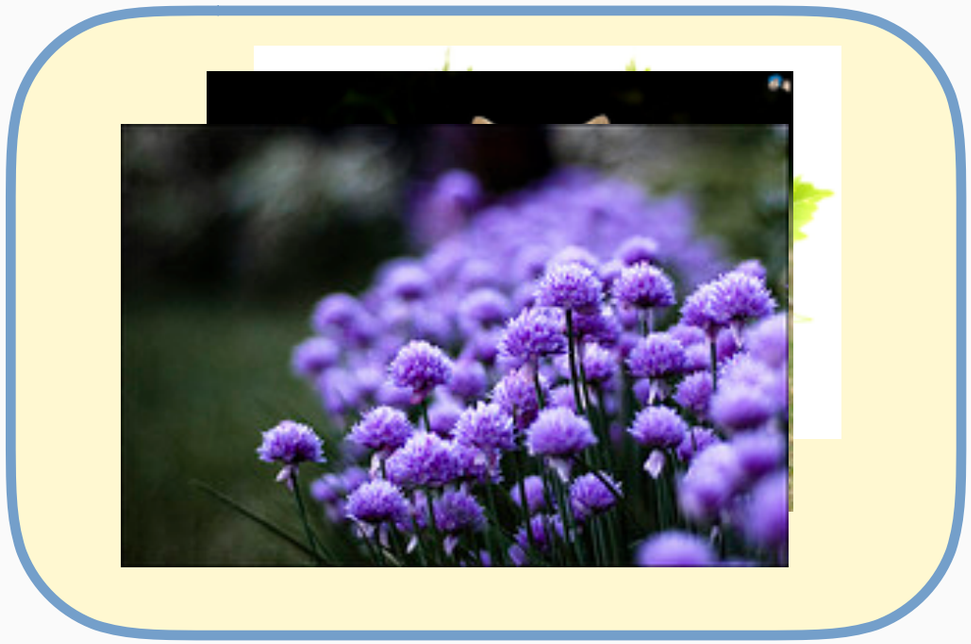
Garden!

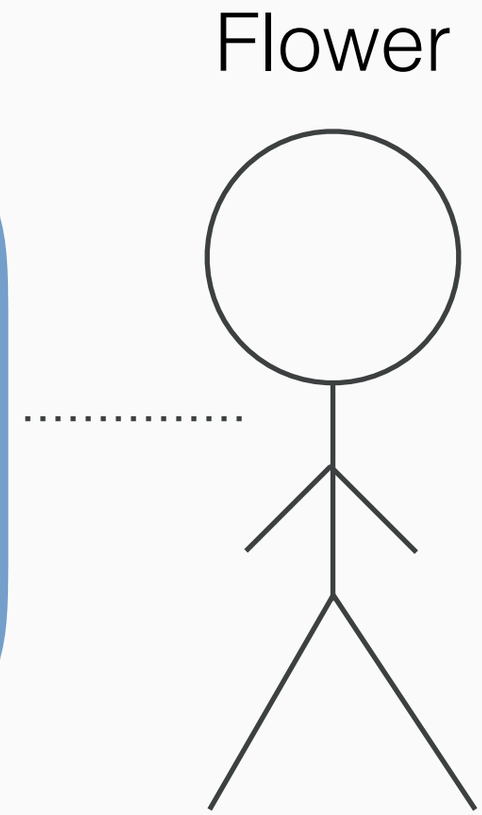
Crowdsourcing



Crowdsourcing

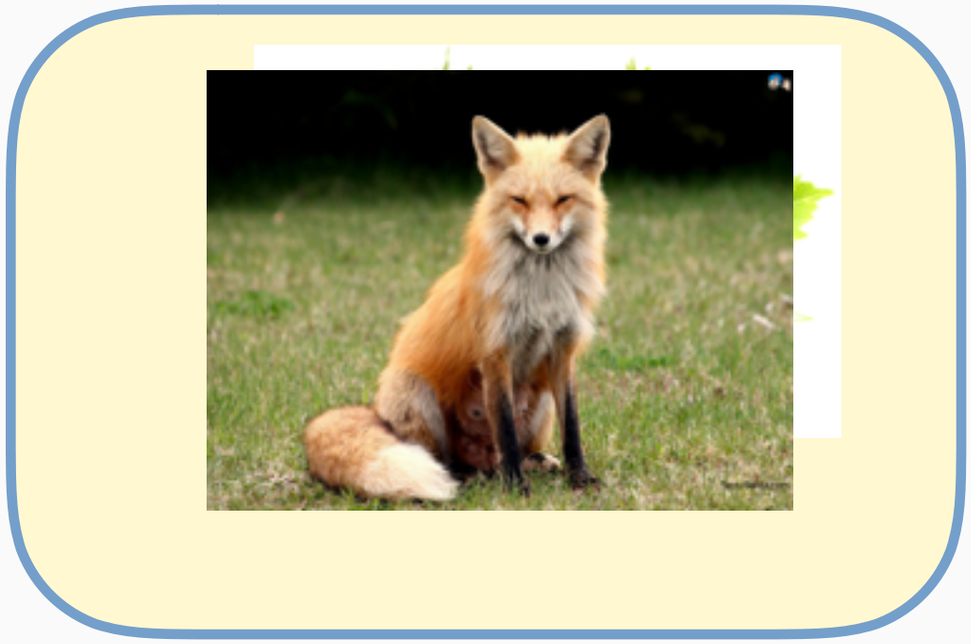
Image	Label

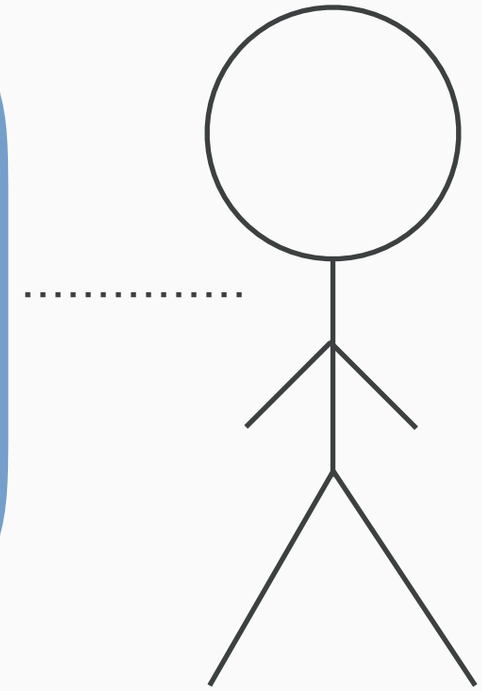




Crowdsourcing

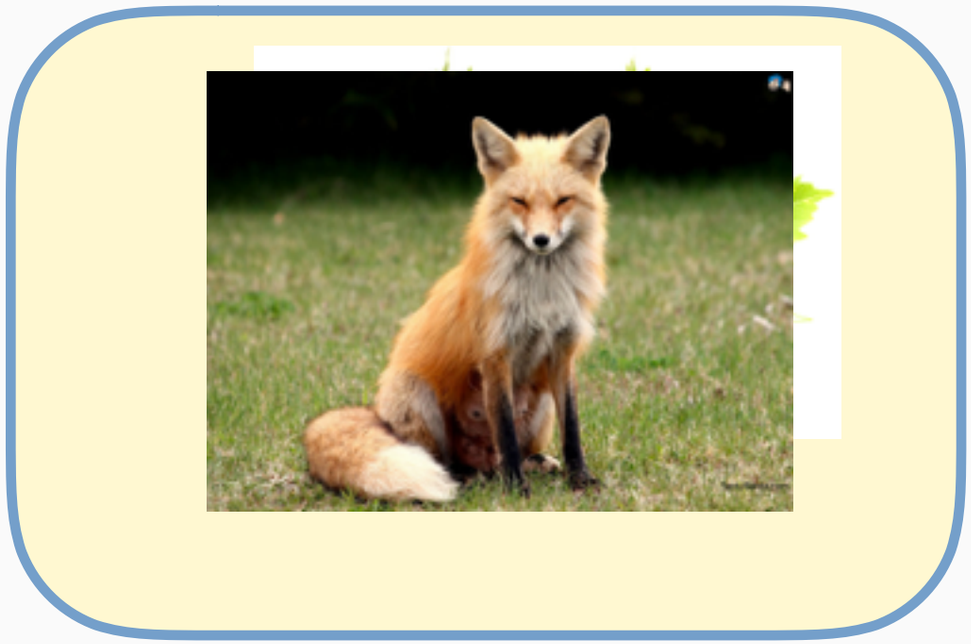
Image	Label
	Flower



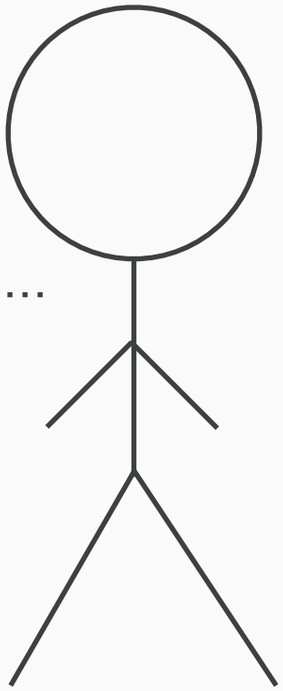


Crowdsourcing

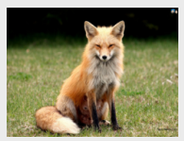
Image	Label
	Flower

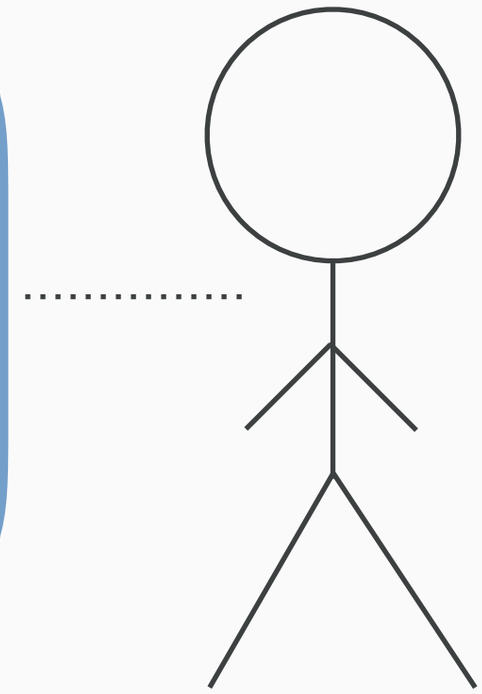


Fox

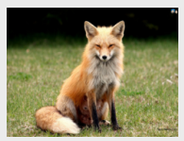


Crowdsourcing

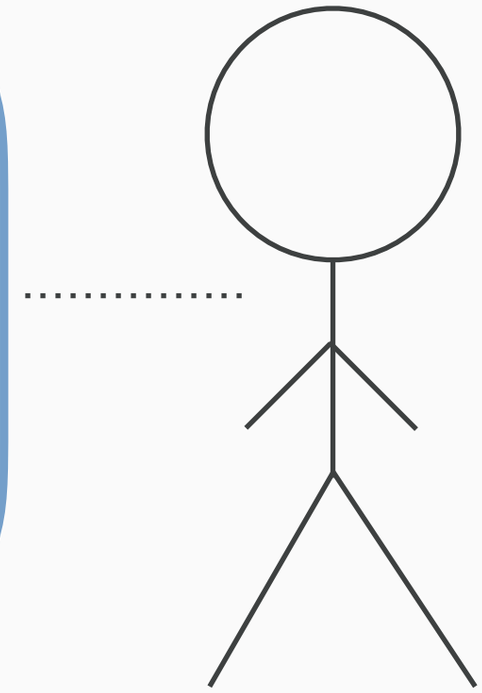
Image	Label
	Flower
	Fox



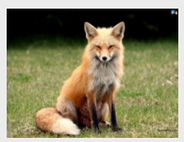
Crowdsourcing

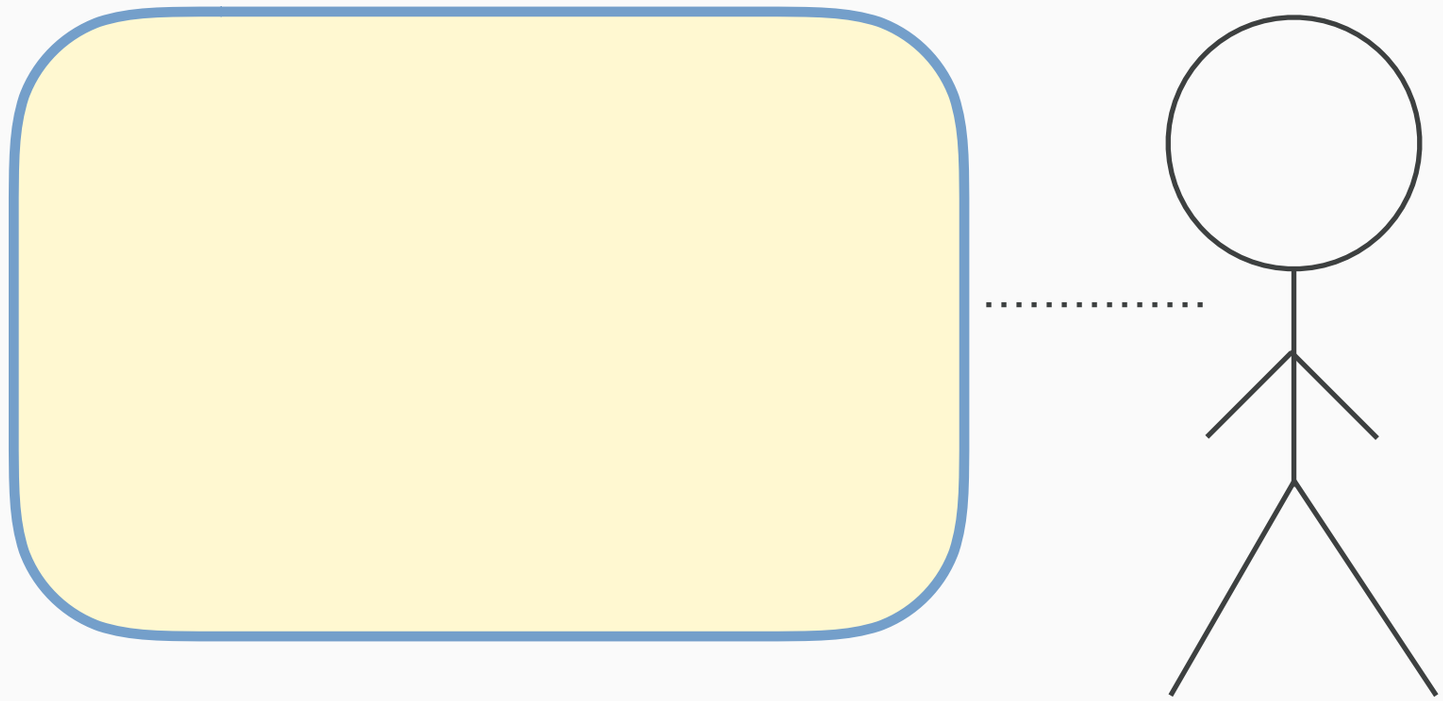
Image	Label
	Flower
	Fox

Grapes

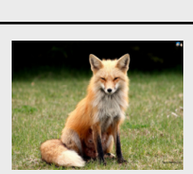


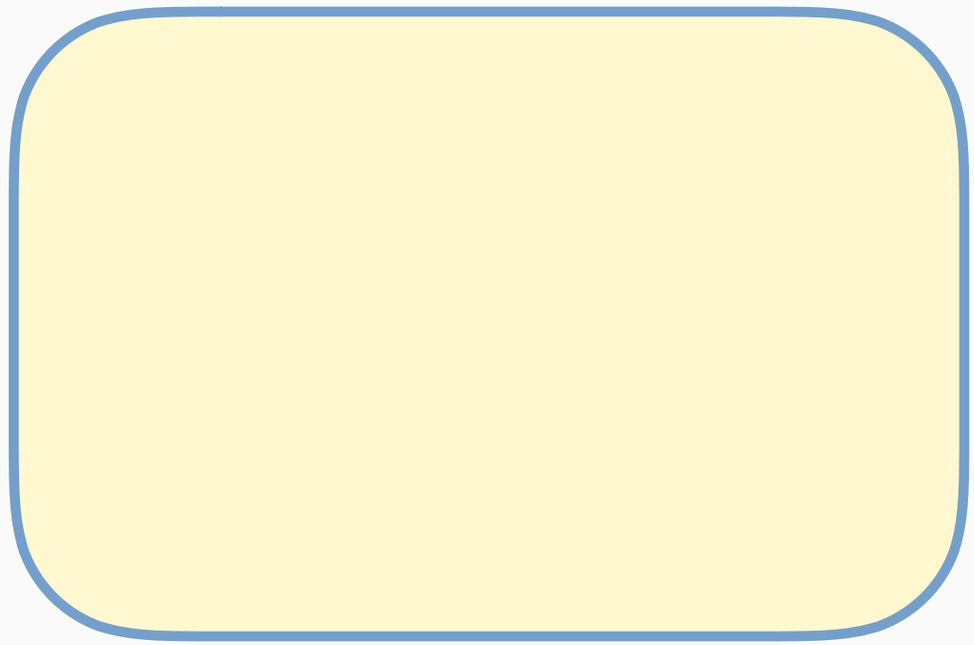
Crowdsourcing

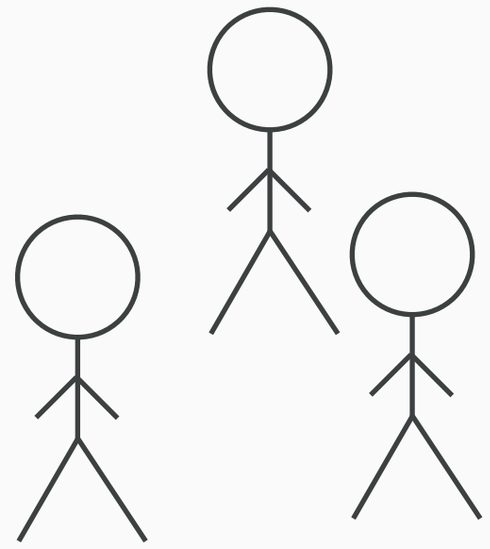
Image	Label
	Flower
	Fox
	Grapes



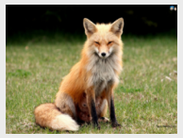
Crowdsourcing

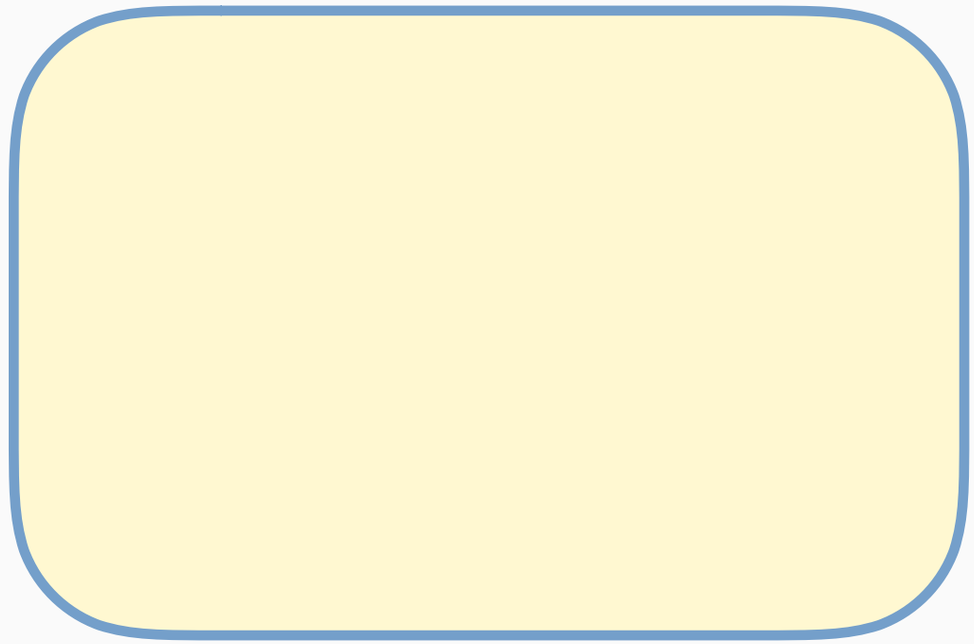
Image	Label
	Flower
	Fox
	Grapes

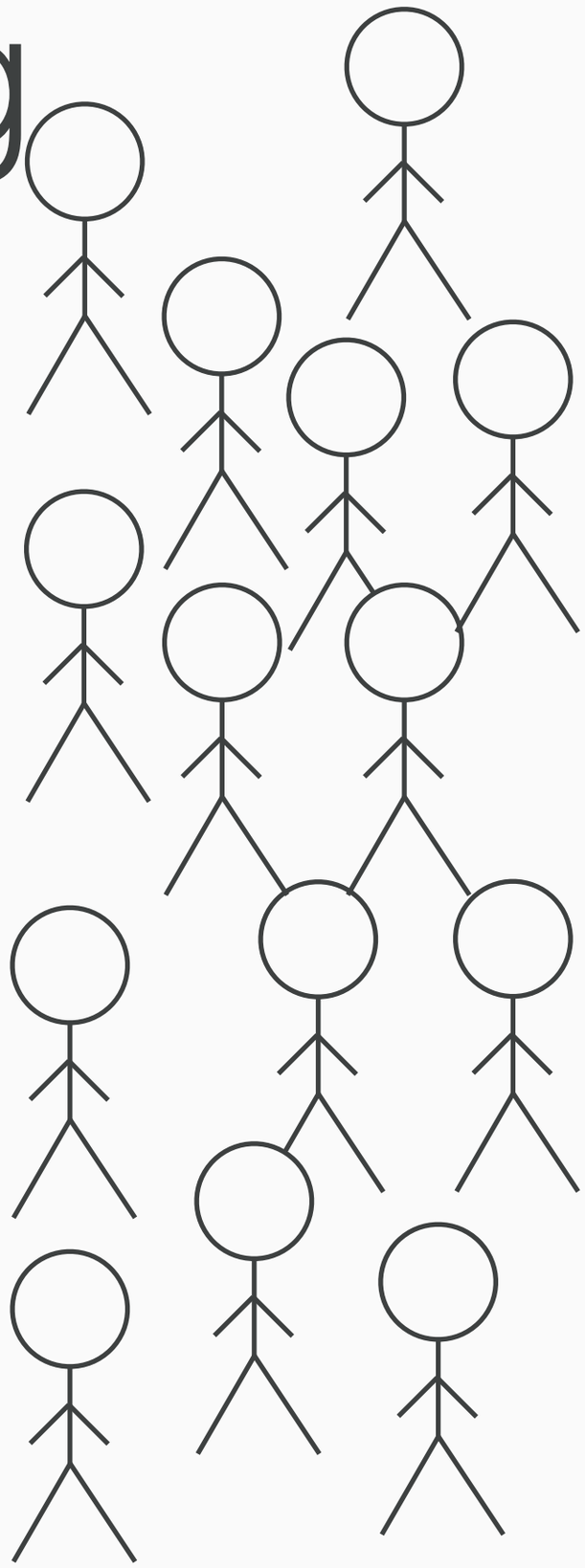




Crowdsourcing

Image	Label
	Flower
	Fox
	Grapes





Crowdsourcing

Amazon Mechanical Turk

Crowdsourcing

Amazon Mechanical Turk

Foldit