

# Case-based Recommendation

Peter Brusilovsky  
with slides of Danielle Lee

# Modern E-Commerce Site

The screenshot shows the CNET Reviews website interface. At the top, there is a navigation bar with the CNET logo, a search bar, and menu items for Reviews, News, Download, CNET TV, How To, and Marketplace. Below the navigation bar, the breadcrumb trail reads "CNET > Reviews > Digital Cameras". The main heading is "Digital Cameras".

The featured product is the Nikon D7000, labeled as a "TOP DIGITAL CAMERA". The product image shows the camera body. Below the image, the text reads "Nikon D7000" followed by a five-star rating (four stars are filled, one is empty) and the price range "\$1,179.95 - \$2,179.95".

To the right of the product listing is a "Find a digital camera" filter section. It includes a search bar with the text "See all digital cameras" and a magnifying glass icon. Below the search bar is a table with four columns: Price, Manufacturer, Zoom range, and Other. The table lists various filters for each category.

Price	Manufacturer	Zoom range	Other
Less than \$100	Sony	Less than 3X	Flash memory
\$100 - \$200	Nikon	3X to 4X	Digital camera type
\$200 - \$300	Panasonic	4X to 8X	Resolution
\$300 - \$400	Canon	8X to 12X	Maximum ISO
\$400 - \$500	Olympus	More than 12X	Weight
\$500 - \$1,200	Fujifilm	<b>See all zoom ranges</b>	Optical sensor type
<b>See all prices</b>	<b>See all manufacturers</b>		

# The Power of Metadata

- Modern e-commerce sites have a range of metadata for each item
  - Travel information presented in its price, duration, accommodation, location, mode of transport, etc.
  - Job information presented in the job kinds, salary, business category of each company, educational level, experience, location etc.
- This data is used in modern Faceted Search, more powerful than keyword search
- The power of metadata can be also used for better recommendation this is the essence of case-based way

The screenshot shows a search results page for 'Digital cameras'. At the top, there's a 'Refine your results' section with four columns of filters: 'Flash memory' (SDXC Memory Card (7)), 'Maximum ISO' (75 or more (7), 400 or more (7), 1250 or more (7), 1600 or more (7), 2500 or more (7), 3200 or more (7)), 'Weight' (4 oz to 8 oz (7), More than 1 lb (1)), and 'More' (Optical sensor type, Zoom range). Below this, 'You selected:' shows filters for 'Nikon', '\$300 - \$400', '12 megapixels', 'Compact', and 'remove all'. The results section shows '8 results' with a 'Show 10 results per page' and 'Sort by: Review date' dropdown. Two products are visible: 'Nikon Coolpix P300 (Black)' with an editor's rating of 4 stars, a user rating of 4 stars, and a price of \$329; and 'Nikon Coolpix S9100 (Red)' with an editor's rating of 4 stars, a user rating of 4 stars, and a price range of \$326 to \$420. Each product listing includes a 'Bottom Line' summary, detailed specifications, and options for 'Price History', 'Set Price Alert', and 'Add to my list'.

## Metadata Could be Used in a Smarter Way

- “6 mega-pixel digital SLR for under \$200”
  - No result is returned → System slavishly respects customers’ queries (“stonewalling”)
- “Another camera like this one but with more optical zoom and a lower price”
  - Too complex for customers to provide this form of feedback directly to the system.
- “I never accepted the cameras above \$1000”
  - Few commercial system to remember customers’ preferences over time.
  - Customers start their search from scratch in every visit.



# Case-Based Recommendation?

- A special form of *content-based* recommendation
- Assumes structured item information with a well defined set of features and feature values.
- Information are represented as a *case* and the system recommends the cases that are *most similar* to a user's preference
- Case-based representation also supports more advanced recommendation dialogues and explanations

# Case-based Reasoning

- Case-based recommendation origins in Case-Based Reasoning (CBR).
  - It is to solve new problems by reusing the solutions to problems that have been previously solved and stored as cases in a case-base.
  - Each case consists of a specification part, which describes the problem and a solution part, which describes the solution of the problem.
    - Solutions to similar prior problems are a useful starting point for new problem solving.
- **“The users would like the similar one that they liked before.”**

# Simple Example of Case-based Recommendation

I want laptop having  
250GB HDD, 1GB  
memory and 14 inch  
screen for \$400



## Product #1

- HDD : 250 GB
- Memory : 2 GB
- Screen Size : 15 inch
- Price : \$550

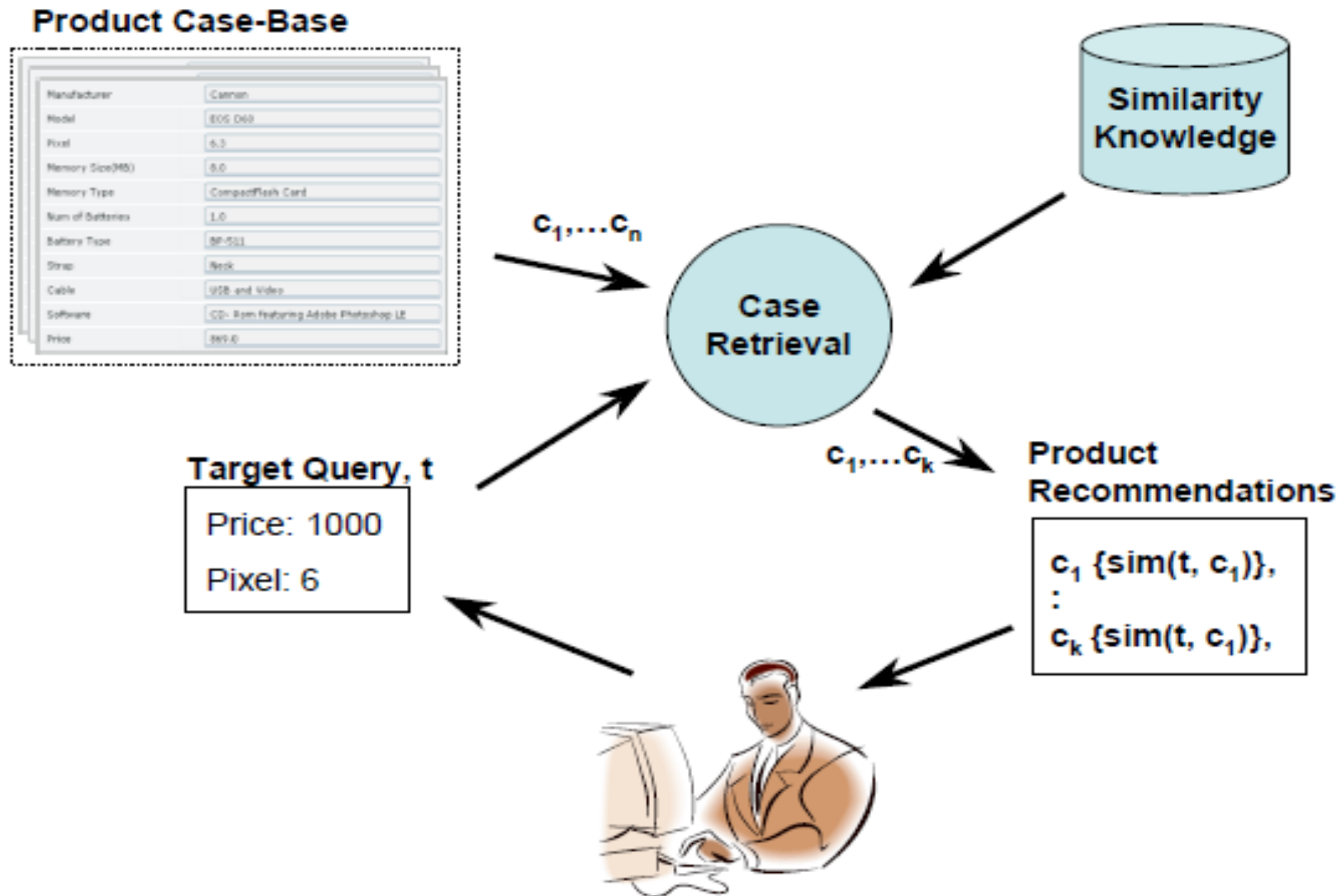
## Product #2

- HDD : 150 GB
- Memory : 1 GB
- Screen Size : 15 inch
- Price : \$450

## Product #3

- HDD : 250 GB
- Memory : 1 GB
- Screen Size : 14.2 inch
- Price : \$420

# Case-based Recommendation





# Case Representation

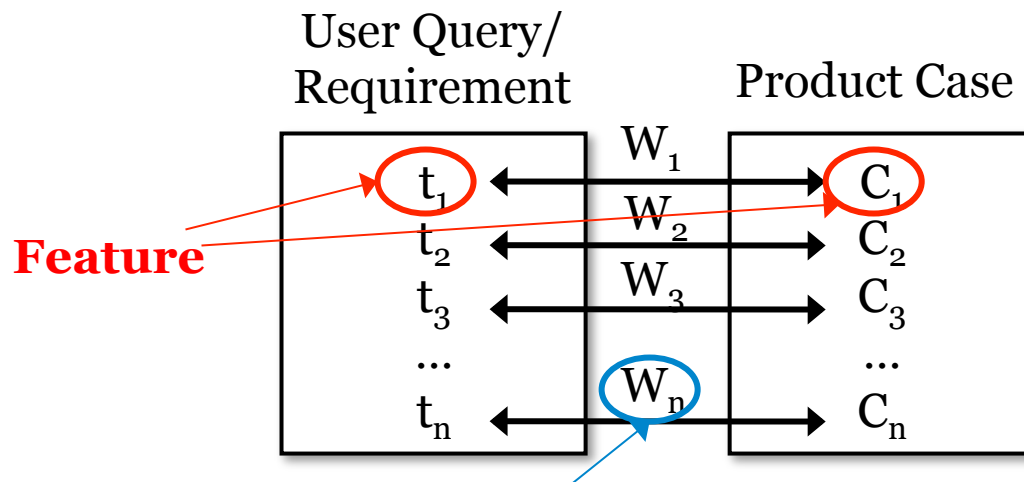
Manufacturer	Cannon
Model	EOS D60
Pixel	6.3
Memory Size(MB)	8.0
Memory Type	CompactFlash Card
Num of Batteries	1.0
Battery Type	BP-511
Strap	Neck
Cable	USB and Video
Software	CD- Rom featuring Adobe Photoshop LE
Price	869.0

Nominal Feature

Numeric Feature

# Similarity Assessment (1)

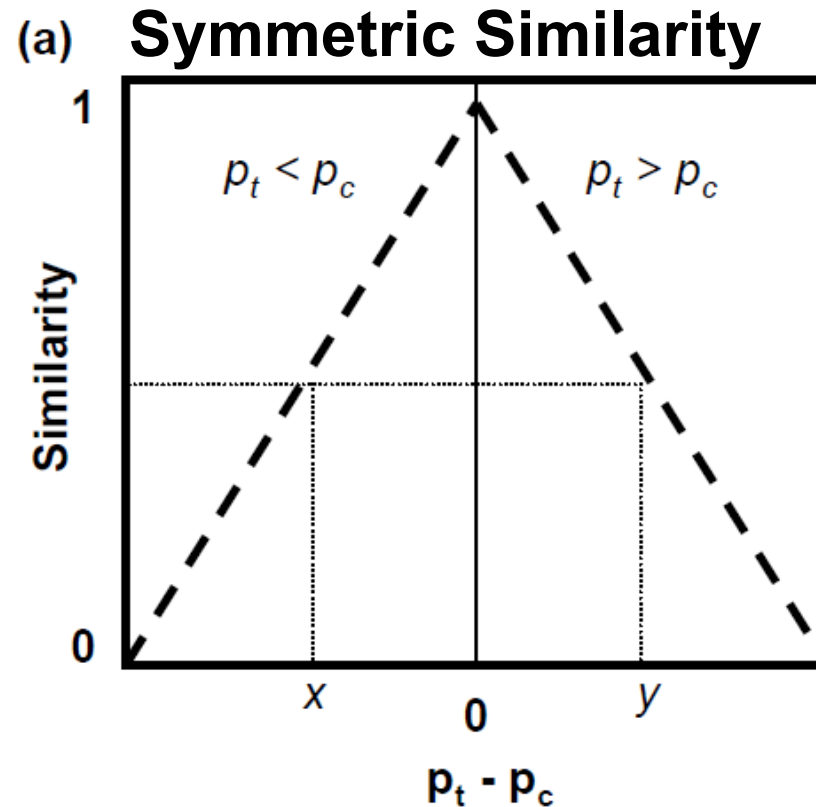
- Similarity metrics that are based on an explicit mapping of case features and the availability of specialized feature level similarity knowledge.



**Relative Importance of the feature**

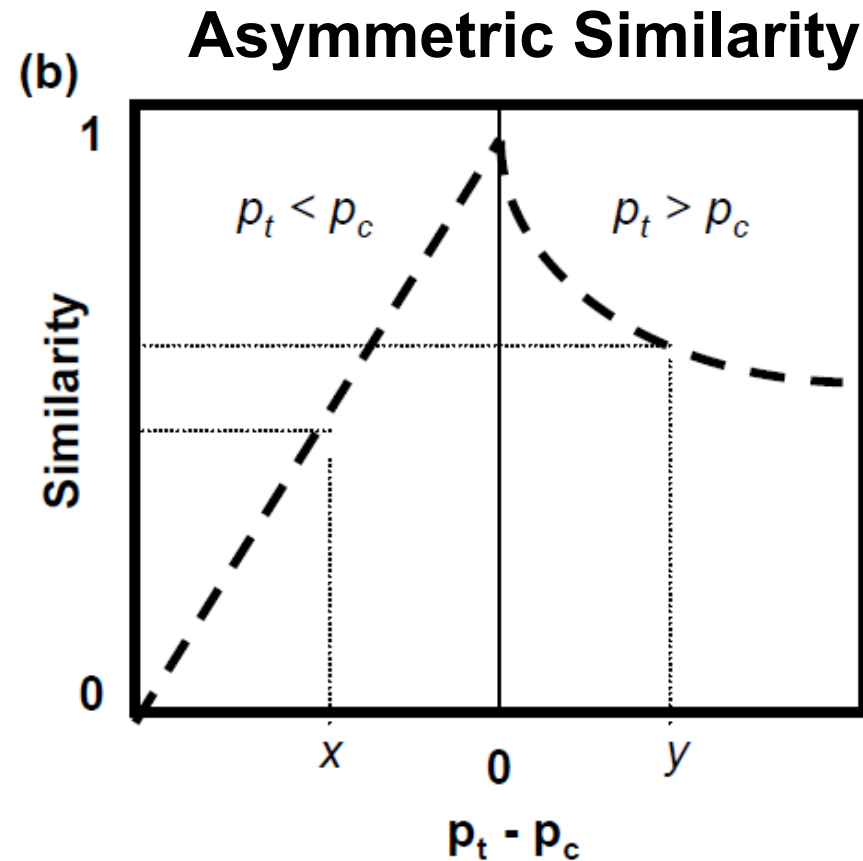
$$\text{Similarity}(t, c) = \frac{\sum_{i=1..n} w_i * \text{sim}_i(t_i, c_i)}{\sum_{i=1..n} w_i}$$

# Similarity Assessment (2)



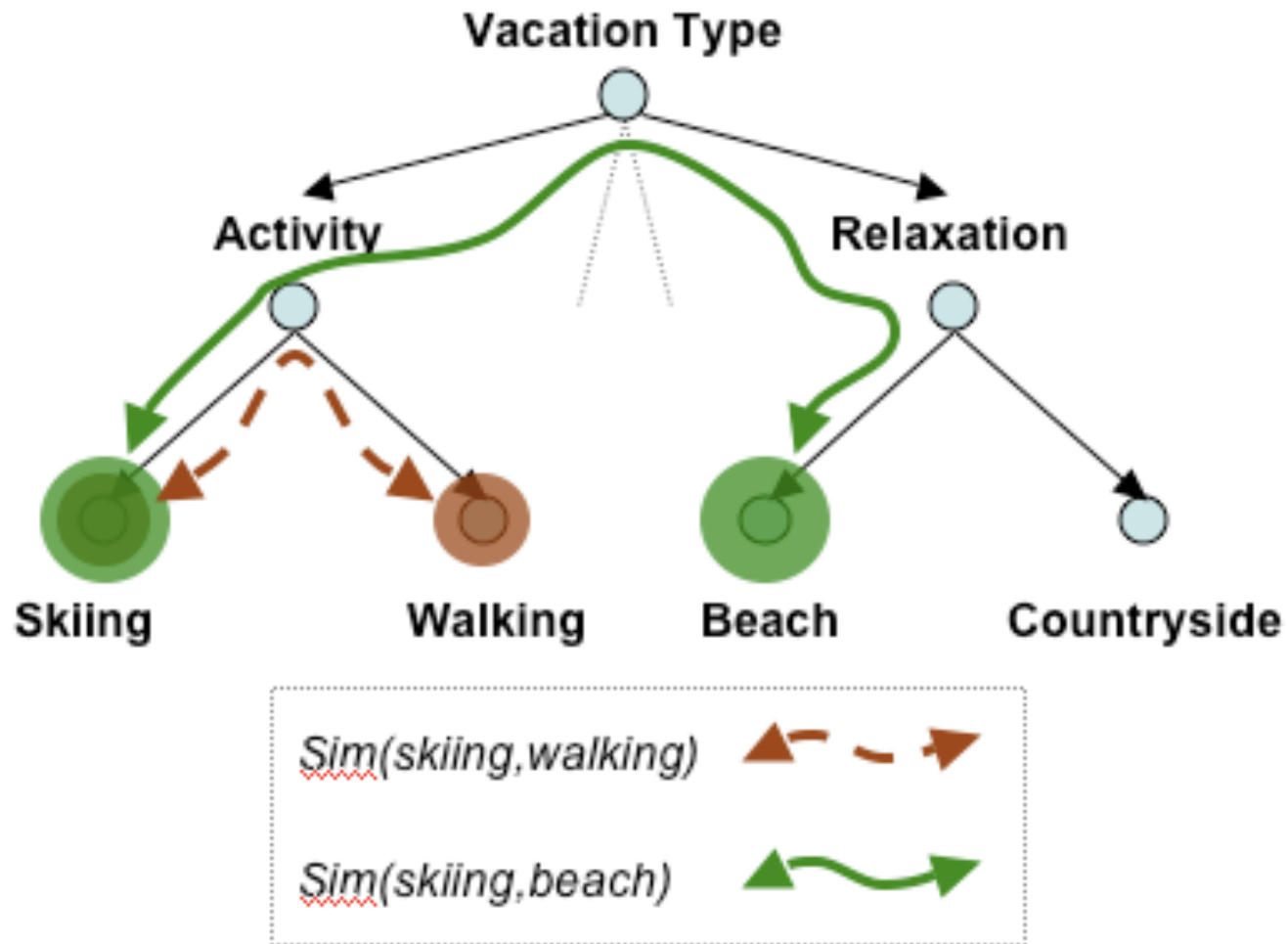
- In symmetric similarity, maximum similarity is achieved when a feature of a candidate case matches that of the target query. **No bias in favor** of either higher or lower values of the corresponding feature.

# Similarity Assessment (3)



- In asymmetric similarity, there is a **bias** to either higher or lower values (i.e. a product that is \$50 cheaper is better than \$50 more expensive)

# Similarity Assessment of Nominal Values



**Partial Ontology of Vacation types**

# Representing Similarity Knowledge

	Small Car	Medium Car	Large Car	SUV	Mini Van
SmCar	1	???	???	???	???
MdCar	???	1	???	???	???
LgCar	???	???	1	???	???
SUV	???	???	???	1	???
Minivan	???	???	???	???	1

# Acquiring Similarity Knowledge

- Based on knowledge made by a domain knowledge expert.
  - Normally it is hand-coded and expensive.
- Machine learning techniques.
  - Using several weight-learning algorithm, even knowledge-poor techniques can result in significant improvements in case-based classification tasks.
- Similarity assessment by users
  - A ‘similarity teacher’ evaluates the ordering for the given set of retrieval results.
  - The selections could be used not only for assessing the similarity but for acquiring users preference.

# Case-based Job Recommendation

Database Developer  
job for a finance-  
related company in  
Boston



**Job #1**  
Database Analyst job for  
Company A

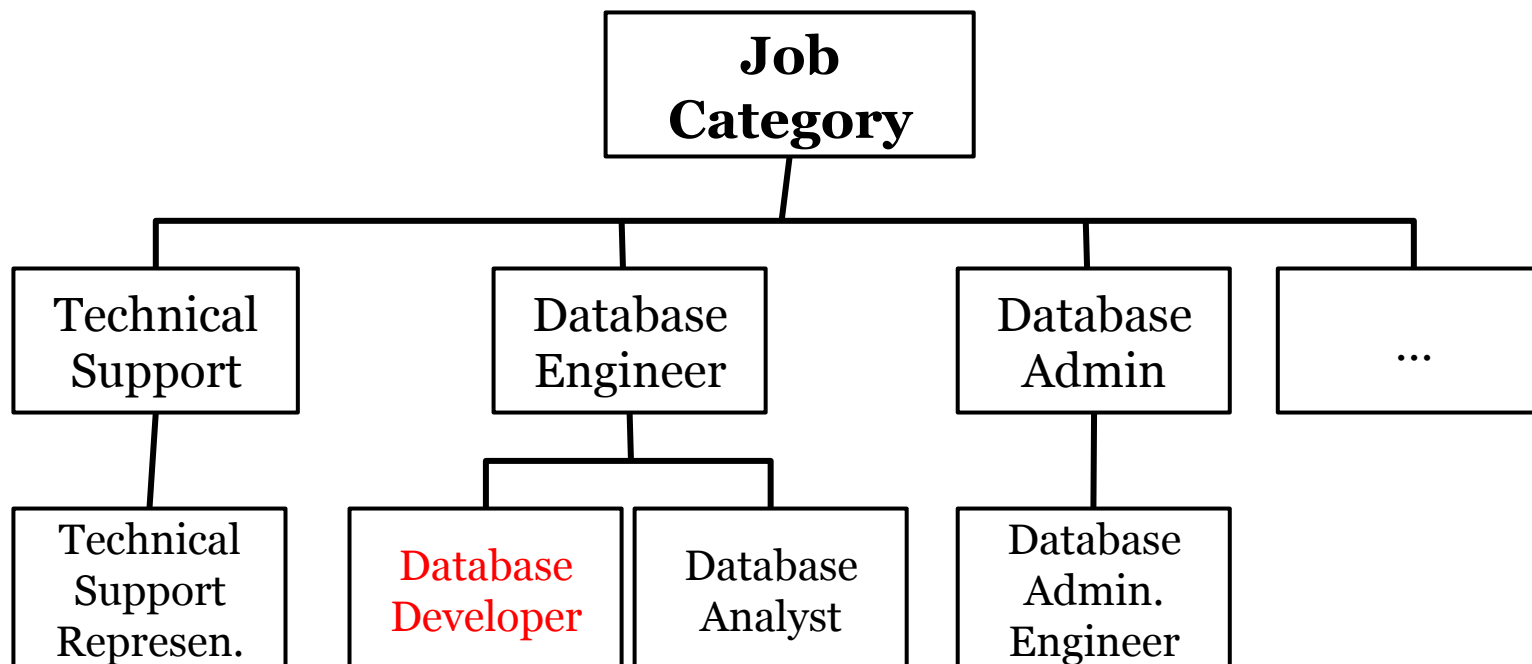
**Job #2**  
Database Administrator job  
for Company B

**Job #3**  
Technical Support Engineer  
for Company C



# Job Related Knowledge (1)

- Partial Ontology about job category.





## Job Related Knowledge (2)

- Taxonomy about Company
  - Company A : Insurance company, downtown in Boston.
  - Company B : Pharmaceutical company, 5 miles distance from Boston.
  - Company C : Computer manufacturing domain, 1.5 miles distance from Boston.

# Proactive - Job Recommendation System

## Proactive

[\[Preference\]](#) [\[Feedback\]](#) [\[Help\]](#) [\[Logout\]](#)

Welcome to Proactive, danielle

[\[Most Recent Jobs\]](#) [\[Matched to Preference\]](#) [\[Recommended Jobs\]](#) [\[My Saved Jobs\]](#) [\[Advanced Search\]](#)

**:: Matched jobs to your preference within 3days ::**  
 Total: 50 (1/3) [\[Preference\]](#) 1 2 3

Job Category	Title	Company	City	State	Position Type	Salary	Experience	Education	Post Date	Relevance
Database Manager	Marketing Database Manager <a href="#">[SAVE]</a>	Info Technologies, Inc.	New York	NY	Full-Time, Contract	unspecified	5-10 Years Experience	unspecified	01-23-08	****
Database Architect	3 NEW TECH OPPORTUNITIES! - .NET DVELOPER (C++, C#), DATABASE ARCHITECT, .NET DEVELOPER MANAGER <a href="#">[SAVE]</a>	Concepts in Staffing	New York	NY	Full-Time, Employee	unspecified	5-10 Years Experience	Bachelor's degree	01-23-08	****
Database Architect	Database Architect <a href="#">[SAVE]</a>	Spot Runner	Los Angeles	CA	Full-Time, Employee	unspecified	10-15 Years Experience	Bachelor's degree	01-23-08	****
Database Architect	Solutions Architect   Senior Solutions Architect <a href="#">[SAVE]</a>	Silver Key	San Francisco	CA	Full-Time, Employee	unspecified	2-5 Years Experience	Bachelor of Science	01-23-08	****
Database Architect	ETL Architect <a href="#">[SAVE]</a>	Kinetic Networks, Inc.	San Francisco	CA	Full-Time, Employee	unspecified	2-5 Years Experience	unspecified	01-23-08	****
Database Specialist	EDUCATIONAL SALES SPECIALIST - EAST <a href="#">[SAVE]</a>	Sika Sarnafil, Inc.	New York	NY	Full-Time, Employee	unspecified	5-10 Years Experience	Bachelor of Science	01-23-08	****
Database Architect	MySQL DBA / Architect <a href="#">[SAVE]</a>	Slide, Inc	San Francisco	CA	Full-Time, Employee	unspecified	2-5 Years Experience	unspecified	01-23-08	****
Database Architect	SQL DAtabase Architect <a href="#">[SAVE]</a>	Global Technical Talent	New York City	NY	Full-Time, Contract	unspecified	2-5 Years Experience	unspecified	01-23-08	****
Database Manager	Release Manager <a href="#">[SAVE]</a>	Kaiser Permanente	Oakland	CA	Full-Time, Employee	unspecified	5-10 Years Experience	unspecified	01-23-08	***
Database Architect	Sr. Database Architect <a href="#">[SAVE]</a>	Careers on the Move	New York	NY	Full-Time, Employee	\$130,000 to \$150,000 per year	10-15 Years Experience	Master of Science	01-23-08	***
Database Developer	Java Developer - Database Developer <a href="#">[SAVE]</a>	Adobe Systems	San Francisco	CA	Full-Time, Employee	unspecified	5-10 Years Experience	Bachelor of Science	01-21-08	***
Technical Support Consultant	Patient Monitoring Technical Consultant <a href="#">[SAVE]</a>	Philips North America	New York	NY	Full-Time, Employee	unspecified	2-5 Years Experience	unspecified	01-20-08	***
Technical Support Consultant	Summit System Support Consultant <a href="#">[SAVE]</a>	Solomon-Page Group	New York	NY	Full-Time, Contract	\$450 to \$650 per year	5-10 Years Experience	Bachelor's degree	01-20-08	***
		BARNES & NOBLE			Full-Time		0-1 Years	Bachelor of		***

## Diversity

“I want a 2-week vacation for two in the sun, costing less than \$750, within 3 hours flying time of Ireland. I expect good night-life and recreation facilities on-site”

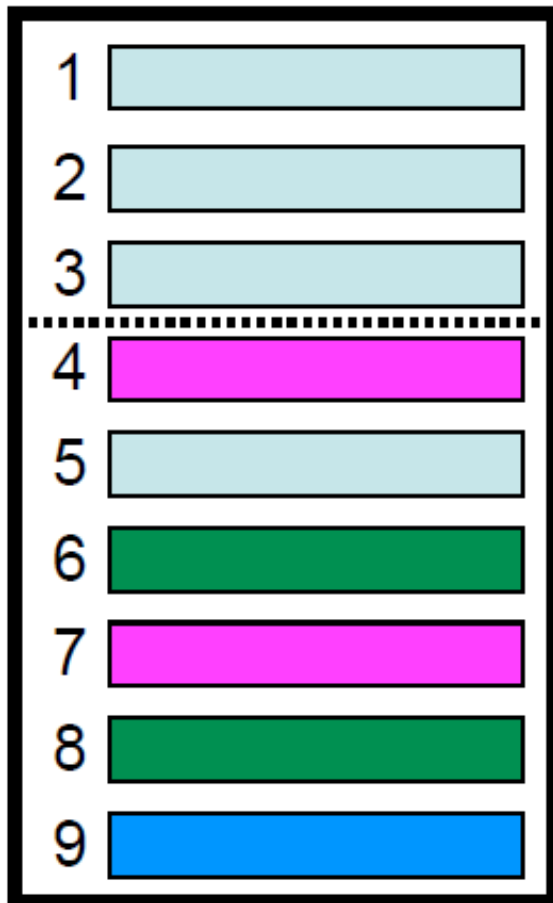


### System suggests ...

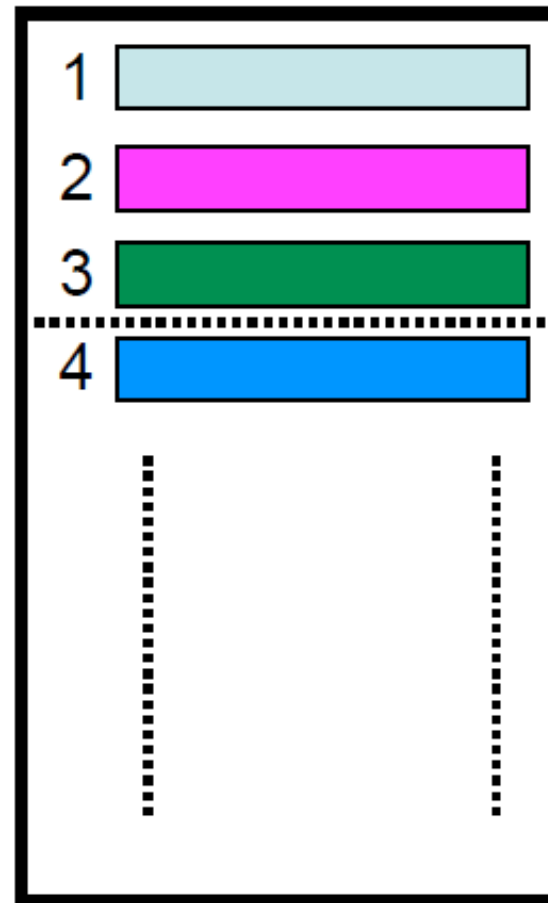
1. Hercules Complex in the Costa Del Sol, Spain on the first two weeks of July
2. Hercules Complex in the Costa Del Sol, Spain on the first two weeks of August
3. Pleasure Complex in the Costa Del Sol, Spain on the last two weeks of July
4. Hercules Complex in the Costa Del Sol, Spain on the last two weeks of July
5. ...

# Similarity vs. Diversity (1)

(a) Similarity-Based



(b) Diversity-Based



## Similarity vs. Diversity (2)

- Bounded Random Selection: from the top  $bk$  most similar cases to the target query, select  $k$  random cases.
  - The diversity could increase but the similarity could also decrease.
- Bounded Greedy Selection: define the diversity of a set of retrieved cases to be the average dissimilarity between all pairs of these cases.
  - 50% improvement in relative diversity with a minor loss of less than 10% in similarity to the target query.
  - A unit drop in similarity can be traded for almost 3 units of diversity using this method.
  - Increased computational efficiency.

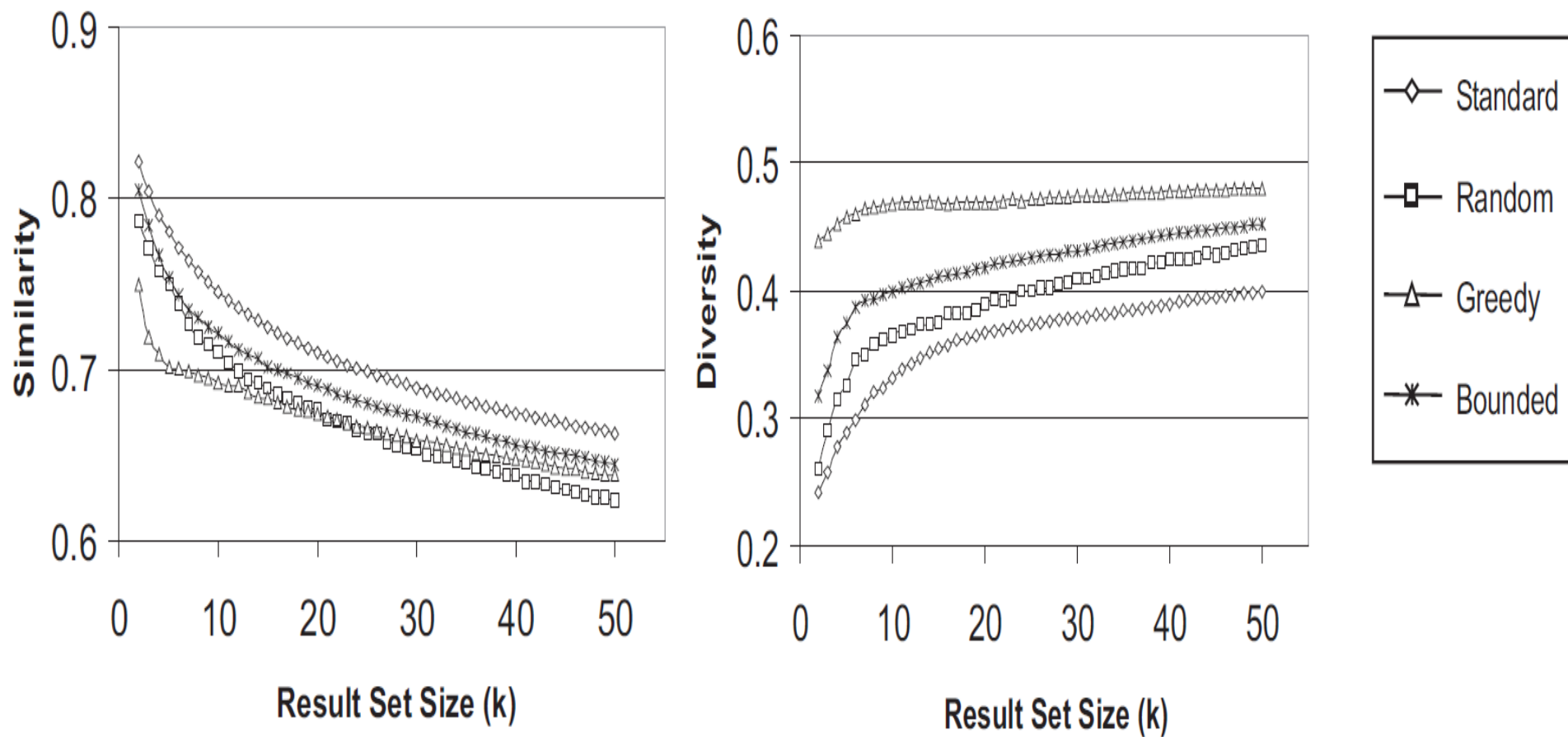
# Bounded Greedy Selection

- Proposed in Smyth & McClave (2001).
- The key idea is the quality metric that combines diversity and similarity.
  1. Select the best  $bk$  cases according to the similarity.
  2. Pick up the one with the highest similarity.
  3. During each subsequent iteration, the case with the highest combination of similarity and diversity with respect to the set of cases selected during the previous iteration.

$$Quality(t, c, R) = Similarity(t, c) * RelDiversity(c, R) \quad (11.4)$$

$$RelDiversity(c, R) = 1 \text{ if } R = \{\}; \\ = \frac{\sum_{i=1..m} (1 - Similarity(c, r_i))}{m}, \text{ otherwise} \quad (11.5)$$

# Similarity vs. Diversity (3)





# Other approaches for diversity

- Shimazu (2001, 2002)
  - It is suitable to the situations where the set of recommended cases is drawn from a set of sufficiently similar cases.
- Similarity layers
  - A set of cases are ranked by their similarity to the target query and partitioned into similarity layers.
  - The lowest similarity layer is selected using an optimal diversity maximizing algorithm.
- Similarity intervals
- Order-based Retrieval
- Compromise-driven approach

# Conversational Recommendation

- Many recommenders operate in a reactive and *single-shot fashion*.
  - If users don't find what they want, the only option is to revise their query and start again.
  - User's initial query could be adapted and refined as they learn more about a particular product-space.
- Conversational recommender is to engage users in an extended, interactive recommendation dialog during the query elicitation.
  - Navigation by asking
  - Navigation by proposing

# Navigation by asking (1) - *Adaptive Place Advisor*

1. **Inquirer:** *Where do you think I should eat tonight?*
2. **Advisor:** What type of food would you like?
3. **Inquirer:** *What types are there?*
4. **Advisor:** You can say things like Chinese, Indian, and Mediterranean.
5. **Inquirer:** *Oh, maybe a cheap Indian place.*
6. **Advisor:** What kind of parking would you like?
7. **Inquirer:** *I don't care, as long as it's in Palo Alto.*
8. **Advisor:** I'm sorry, I don't know of any restaurants like that, would you like to search for any price?
9. **Inquirer:** *No, I think I'd like Chinese instead.*
10. **Advisor:** How does Mandarin Gourmet on 420 Ramona sound?<sup>1</sup>
11. **Inquirer:** *No, what else do you have?*
12. **Advisor:** How about Jing-Jing Szechwan Hunan Gourmet on 443 Emerson Street?
13. **Inquirer:** *Sure, that sounds fine.*



## Navigation by asking (2)

- The selection of questions does matter.
  - Entropy - Dolye & Cunningham (2000)
  - Information gain - Quinlan (1986); Quinlan (1993) and Shimazu (2001); Shimazu (2003)
  - Influence - Schmitt (2002)
- However, directly asking questions to users can be burden to them.
  - A long list of questions, uncertain answers or rejected questions.
  - Significant interfacing burden



## Navigation by proposing (1)

- The users are presented with one of more recommendation alternatives, rather than a question, during each recommendation cycle.
  - Rating based feedback
  - Critique based feedback: Constraints over certain features of recommendations
  - Preference based feedback: Expressed preference for one alternative over the others

# Critique based feedback

UKRAINIAN VILLAGE. TWO bedroom rehab garden apartment. Lr, Eurokitchen, hwfl, excellent security, forced air, lots of closets, laundry in building. Garage space included. Dogs OK. Available immediately. \$600/ mo. 312-489-1554. / ;

Phone: 312-489-1554

2-bedrooms

\$600

60622  
(West Town  
Bucktown)

This apartment is OK, but make it...

bigger

cheaper

nicer

safer

This neighborhood could be more...

convenient

conservative


dynamic

# Compound Critiques


**QWIKSHOP.COM** HOME : ABOUT THIS PROJECT : CONTACT


>> Digital Cameras **Unit Critiques**

Shop for: Digital Cameras Computers Holidays



**Product Found: Canon EOS 30**  
6.3 Megapixel CMOS sensor  
7-point wide-area AF  
High-performance DIGIC processor  
100-1600 ISO speed range  
Compatible with all Canon EF lenses and EX Speedlites  
PictBridge, Canon Direct Print and Bubble Jet Direct compatible - no PC required

**I've found the Camera I want!** 

**No lets start again** 

**Adjust your preferences to find the right camera for you**

Manufacturer	X	Canon	X
Optical Zoom	↓	7x	↑
Memory (MB)	↓	512	↑
Weight (Grams)	↓	780	↑
Resolution	↓	6.2 M Pixels	↑
Size	X	Large	X
Case	X	Magnesium	X
Price	↓	995	↑

**We have more matching cameras with the following:**

- 1. Less Memory and Lower Resolution and Cheaper [EXPLAIN](#) [PICK](#)
- 2. Different Manufacturer and Less Zoom and Lighter [EXPLAIN](#) [PICK](#)
- 3. Lighter and Smaller and Different Case [EXPLAIN](#) [PICK](#)

**Explain:**  
**1. Less Memory and Lower Resolution and Cheaper**

-----

This Critique covers **153** other Digital Cameras

-----

**Less Memory**  
Current Value: 512 MB  
Critique: Less Than  
Remaing: (0 to 256 MB)

**Lower Resolution**  
Current Value: 6.2 M Pixels  
Critique: Less Than  
Remaing: (1.4 to 5.9 M Pixels)

**Cheaper**  
Current Value: 995 €  
Critique: Less Than  
Remaing: (75€to 960€)

[PICK](#)

# Explanations and Clustering (Pu)

## The top candidate according to your preferences

Manufacturer	Price	MegaPixels	Optical zoom	Memory type	Flash memory	LCD screen size	Depth	Weight	
Canon	\$242.00	5.0 MP	3x	CompactFlash Card	32 MB	1.8 in	1.37 in	8.3 oz	<a href="#">choose</a>

## We have more products with the following they are cheaper and lighter, but have fewer megapixels

Nikon	\$167.95	4 MP	3x	SD Memory Card	14 MB	1.8 in	1.4 in	4.6 oz	<a href="#">choose</a>
Canon	\$230.00	4.1 MP	3x	CompactFlash Card	32 MB	1.5 in	1.09 in	6.53 oz	<a href="#">choose</a>
Canon	\$180.00	3.3 MP	3x	SD Memory Card	16 MB	2 in	0.83 in	4.06 oz	<a href="#">choose</a>
Canon	\$219.18	4.2 MP	4x	MultiMedia Card	16 MB	1.8 in	1.51 in	6.35 oz	<a href="#">choose</a>
Canon	\$163.50	3.2 MP	4x	MultiMedia Card	16 MB	1.8 in	1.5 in	6.3 oz	<a href="#">choose</a>
Canon	\$199.40	3.2 MP	2.2x	SD Memory Card	16 MB	1.5 in	1.4 in	5.8 oz	<a href="#">choose</a>

## they have more megapixels and bigger screens, but are more expensive

Sony	\$365.00	7.2 MP	3x	Internal Memory	32 MB	2.5 in	1.5 in	6.9 oz	<a href="#">choose</a>
Canon	\$439.99	7.1 MP	3x	SD Memory Card	32 MB	2 in	1.04 in	6 oz	<a href="#">choose</a>
Fuji	\$253.00	6.3 MP	4x	XD-Picture Card	16 MB	2 in	1.4 in	7.1 oz	<a href="#">choose</a>
Sony	\$336.00	7.2 MP	3x	Internal Memory	32 MB	2 in	1 in	5 oz	<a href="#">choose</a>
Nikon	\$304.18	7.1 MP	3x	Internal Memory	13.5 MB	2 in	1.4 in	5.3 oz	<a href="#">choose</a>
Olympus	\$334.00	7.4 MP	5x	XD-Picture Card	32 MB	2.0 in	1.7 in	7.1 oz	<a href="#">choose</a>

## they are lighter and thinner, but have less flash memory

Pentax	\$238.99	5.3 MP	3x	Internal Memory	10 MB	1.8 in	0.8 in	3.7 oz	<a href="#">choose</a>
Canon	\$273.18	4.0 MP	3x	SD Memory Card	16 MB	2 in	0.82 in	4.59 oz	<a href="#">choose</a>
Nikon	\$329.95	5.1 MP	3x	Internal Memory	12 MB	2.5 in	0.8 in	4.2 oz	<a href="#">choose</a>
Canon	\$316.18	5.3 MP	3x	SD Memory Card	16 MB	2 in	0.81 in	4.59 oz	<a href="#">choose</a>
Casio	\$386.00	7.2 MP	3x	Internal Memory	8.3 MB	2.5 in	0.88 in	4.48 oz	<a href="#">choose</a>
Fuji	\$309.18	6.3 MP	3x	XD-Picture Card	16 MB	2.5 in	1.1 in	5.5 oz	<a href="#">choose</a>





# Case-based User Profiling (1)

- Conversational recommenders can react to the feedback provided by users within each session.
  - In-session personalization only and two users who respond in the same way within a session will receive the same recommendations.
  - How can the systems adapt to the users' persistent preference?
- It is important for the recommenders to learn and maintain a long-term model of a user's recommendation preferences.



## Case-based User Profiling (2)

- CB leverages available content descriptions of cases as a form of case-based user profile.
  - User profile is made of a set of cases and the preference (like or dislike)
- CASPER : Online recruitment system using implicit user profile (from positive and negative points of view) and this profile is used to re-order the recommendations.
  - The Personal Travel Assistant also has similar approach.



# Feature Level User Profiling

- The preference related to features and their values such as preferred values for a particular features, the relative importance of a particular attributes, etc.
  - In restaurant recommendation, the kind of cuisine has an importance weight of 0.4 and parking facilities have a preference weight of 0.1. The user also prefers Italian cuisine with 0.35 weight to German food with 0.1 weight.



## Hybridization of CB and CF - PTV

- O' Sullivan, et al. (2002)
- To solve sparsity problem or latency problem in CF, case-based technology was used.
- By the derived similarity knowledge using data mining technology, the relationships between information items was extended.
- Increased recommendation coverage and recommendation accuracy.



Question?