

## Introduction to Content-Based Recommenders

### Basic Idea: Stable Preferences (measured by content attributes)

- Let's consider some examples:
  - News – I prefer stories on technology, University of Minnesota, Minnesota Vikings, restaurant reviews
  - Clothing – I prefer cotton, blue, low-priced, casual
  - Movies – I prefer Tom Hanks, Sandra Bullock, Woody Allen, Comedy
  - Hotels – I prefer 24-hour front desk, room service, internet, pool

## The key ideas

- Model items according to relevant attributes
- Model or reveal user preferences by attribute
- Voila! A Recommender

## Learning Objectives (I)

After completing this module, you should be able to:

- To understand the range and value of content-based approaches to recommendation
  - Pure information filtering systems
  - Case-based reasoning systems
  - Knowledge-based navigation systems
- Identify situations in which content-based filtering does or does not work well
- Build vector models representing user preferences and item expression in terms of keywords, tags, or other attributes

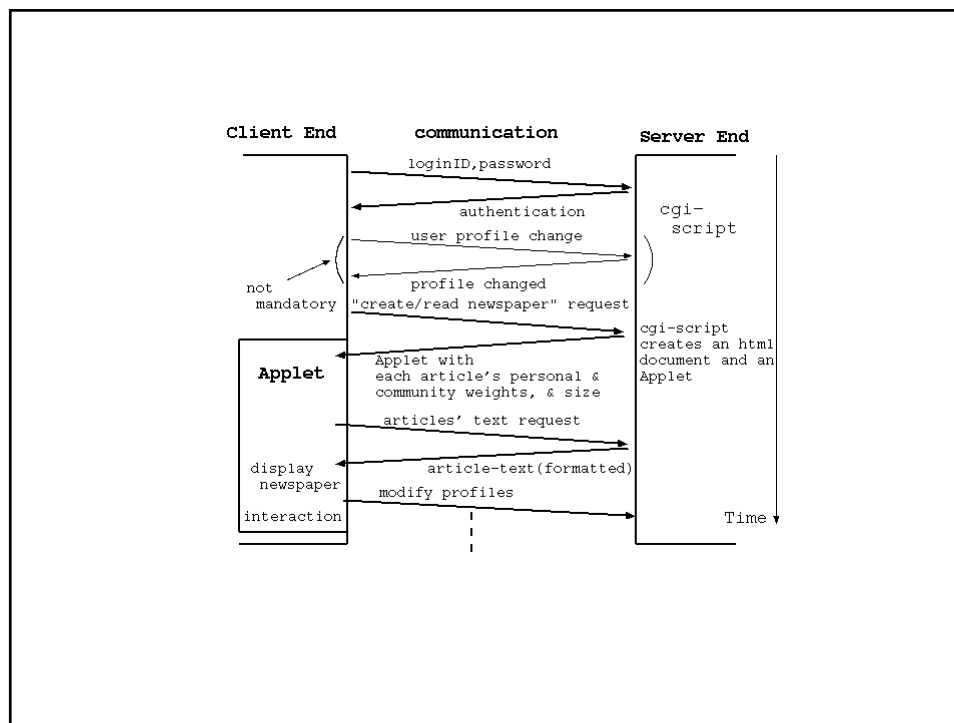
## Learning Objectives (II)

After completing this module, you should be able to:

- Compute recommendations and predictions using TF-IDF and cosine similarity algorithms
- Understand, explain, and apply data normalization in content-based filtering
- Understand and be able to explain the TF-IDF algorithm, and in particular the potential problem that the IDF weighting component is intended to fix
- Complete a programming project using the LensKit toolkit -- specifically the implementation and customization of tag-based content-based recommendation (Honors Track)

## Content-Based Filtering

- Key concept: building a vector of attribute or keyword preferences
- Example: Krakatoa Chronicle  
<http://www.w3.org/Conferences/WWW4/Papers/93>  
Kamba, Bharat, and Albers (WWW '95)



## Wide range of Possibilities

- User could build own profile (awkward)
  - But allowing user to edit a profile can be valuable
- Infer profile from user actions
  - Read, Buy, Click
- Infer profile from explicit user ratings
  - How to map from item preference to attribute preference
- We merge actions/explicit into infer from ratings (explicit and implicit)

## How to build preferences?

- Let's start with the idea of a set of "keywords" that users may like, dislike, or not have an opinion on
- We could simply count the number of times the user chooses (or fails to choose) items with each keyword
- Or we can get more sophisticated (and we will!)

## How to use preferences

- Given a vector of keyword preferences
  - Do we just add up likes and dislikes?
  - Can we figure out which keywords are more and less relevant?
- Forward reference: TFIDF

## Assignments and Assessment

- Our assignments will be based on this model
  - Spreadsheet exercises: building a profile and using it to predict a few cases
  - Programming exercise: building a content-based recommender
- Plus we'll have a module quiz to test comprehension
- But first, a few other approaches ...

## Case-Based Recommendation

- The concept:
- Structure a database of cases around a set of relevant attributes (e.g., camera price, zoom, pixels)
- Query based on an example or attribute query, and retrieve relevant cases
- Open issue: Many ways to structure interaction

## etown's Ask Ida

- No longer exists (old screenshots)
- Uses an interview process to elicit preferences over attributes
- Uses preferences to recommend products
- Uses recommendation as a point to elicit further preferences
- Note: not intended as permanent preferences – just transactional

etown.com™ THE Consumer Electronics Source

Customer Service | Your Account | Shopping Cart | July 17, 2000

Search:

**News & Views**

- Thin Just In**
  - Dynaudio shoes generate electricity
  - Showtime phases in Dolby 5.1 sound
  - EMI to vend pay-per-download songs
  - Streaming A/V on your cell phone?
  - Stephen King plans serial e-book
  - Jung, Hughes team on movie-by-satellite
  - Will Sony Palm unit hit U.S. by Xmas?
- New Reviews/Products**
  - EXCLUSIVE: Toshiba 480p-mst DVD
  - Bush TV center has nice price
  - Vidikron projector: 'an eye opener!'
  - Panamax nixes electrical spikes
  - JVC boombox has stylish looks
  - DBS security recommends AT&T phone
  - EXCLUSIVE: Sharp DVD player
  - EXCLUSIVE: Philips DVD recorder
  - EXCLUSIVE: Outlaw 6.1 A/V receiver
- Features / Columns**
  - Unclear on the THX concept?
  - In the MIs: did we ask for DVD-A?
  - Don't miss our updated FAQ for DTV!
  - Camcorder Corner: Fade to black
  - How to set up your subwoofer

**Browse & Buy**

- Home Theater**
  - NEW! Satellite DVD Receivers
  - Surround Separates Essential
  - Expanded Content Receivers
  - Receivers: EXC
- Television**
  - Quick View TV New Projection TV
  - Quick View TV New LCD TV
  - NEW! Personal TV DTV
- Camcorders**
  - NEW! Digital VHS Camcorders
  - NEW! Personal Digital Camcorders
- Home Audio**
  - NEW! Portable Receivers
  - NEW! Compact Discs
  - NEW! Stereo Separates
  - NEW! Stereo Separates
  - NEW! Stereo Separates
- Portable Tech**
  - NEW! Portable Receivers
  - NEW! Portable Receivers
  - NEW! Portable Receivers
  - NEW! Portable Receivers
- Telecom**
  - NEW! Cordless Phones
  - NEW! Cordless Phones
  - NEW! Cordless Phones
  - NEW! Cordless Phones

**The Top 10 best-selling items:**

etown.com™ THE Consumer Electronics Source

Customer Service | Your Account | Shopping Cart | July 17, 2000

Search:

**Categories**

- Home Theater
- Camcorders
- Home Audio
- Television
- Portable Tech
- Telecom

**News & Views**

- Thin Just In**
  - Dynaudio shoes generate electricity
  - Showtime phases in Dolby 5.1 sound
  - EMI to vend pay-per-download songs
  - Streaming A/V on your cell phone?
  - Stephen King plans serial e-book
  - Jung, Hughes team on movie-by-satellite
  - Will Sony Palm unit hit U.S. by Xmas?
- New Reviews/Products**
  - EXCLUSIVE: Toshiba 480p-mst DVD
  - Bush TV center has nice price
  - Vidikron projector: 'an eye opener!'
  - Panamax nixes electrical spikes
  - JVC boombox has stylish looks
  - DBS security recommends AT&T phone
  - EXCLUSIVE: Sharp DVD player
  - EXCLUSIVE: Philips DVD recorder
  - EXCLUSIVE: Outlaw 6.1 A/V receiver
- Features / Columns**
  - Unclear on the THX concept?
  - In the MIs: did we ask for DVD-A?
  - Don't miss our updated FAQ for DTV!
  - Camcorder Corner: Fade to black
  - How to set up your subwoofer

**Interactive Knowledge Bank**

**Welcome!**

I'm Ida, your interactive decision advisor. I can help you find the products that best suit your needs and preferences. I combine etown.com's expertise on consumer electronics with state-of-the-art artificial-intelligence techniques from Ask Jeeves. I think you'll find shopping for electronics can be easy and fun!

**Select the product category you are interested in:**

- Home Theater**
  - DVD Players
  - A/V Receivers
  - VCRs
- Portable Tech**
  - Digital Cameras
  - Boomboxes
  - Handhelds/gam PCs
- Home Audio**
  - Compact Systems
  - CD Players
- Telecom**
  - Cordless Phones
- Camcorders**
  - 8mm VHS and DV

**NEW EPISODES**

**What is GPS?**

**Interactive Knowledge Bank**

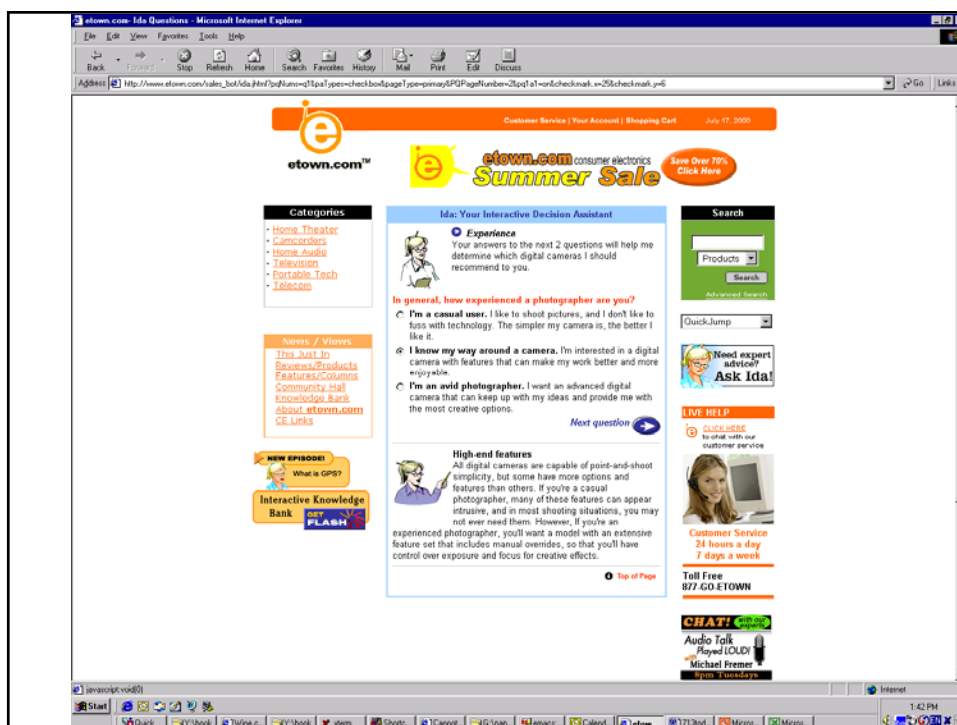
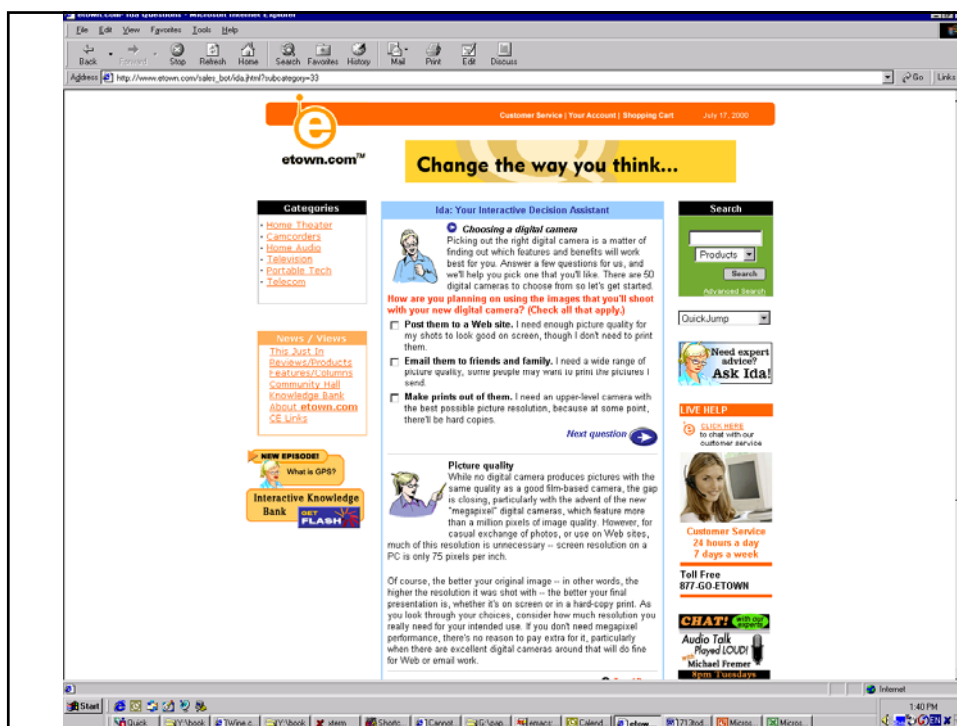
**CHAT: Gchat**

**Audio Talk**

**Michael Premer**

**8pm - 10pm**





etown.com™

(The world's largest Selection) NETFLIX

Categories

- Home Theater
- Camcorders
- Home Audio
- Television
- Portable Tech
- Telecom

News / Views

This Just In

Features/Columns

Community Hall

Knowledge Bank

About etown.com

CE Links

NEW EPISODE!

What is GPS?

Interactive Knowledge Bank

FLASH

Ida: Your Interactive Decision Assistant

**Budget choice**

I have divided all the digital cameras into 4 budget categories. If you would like, I can exclude products that cost more than you are willing to spend.

**About how much would you consider spending on a new digital camera?**

- Up to \$300.** In this price range you can expect to find a fairly basic digital camera with no zoom lens and standard picture resolution (under a megapixel). These cameras will produce images that are fine for posting to the Web or emailing, but do less well when they're printed on paper. All of the 8 digital cameras in this range fit your expressed needs well.
- Up to \$600.** At the lower end of this price range you'll find some megapixel cameras that produce images that can be printed out at small sizes with acceptable results. Toward the high end, there are feature-rich cameras that produce fairly high-quality images that translate into sharp prints at sizes up to 4 x 6 inches. All of the 24 digital cameras in this range fit your expressed needs well.
- Up to \$900.** If bigger-sized hard copies are what you're after, you'll probably have to step up to this price range. Cameras here generally have a zoom lens and a minimum resolution of 1.3 megapixels; some top the 2-megapixel mark. Most models in this category also have more flexibility in their operation, with more manual settings. All of the 46 digital cameras in this range fit your expressed needs well.
- \$1000 or more.** This is "premium" territory – the line between professional and consumer gear. Many models have 2-megapixel or better resolution, as well as plenty of manual options for more advanced users. Zoom is standard, and some models may have options for adding lenses (wide angle or telephoto). All of the 50 digital cameras in this range fit your expressed needs well.

See Initial recommendations

Search

Products

Search

Advanced Search

QuickJump

Need expert advice? Ask Ida!

LIVE HELP

Click Here

to chat with our customer service

Customer Service

24 hours a day

7 days a week

Toll Free

877-GO-ETOWN

CHAT! Click Here

Audio Talk

Played LOUD!

Michael Bremer

Open 1 weekdays

etown.com™

etown.com consumer electronics

Summer Sale

Save Over 70% Click Here

Categories

- Home Theater
- Camcorders
- Home Audio
- Television
- Portable Tech
- Telecom

News / Views

This Just In

Features/Columns

Community Hall

Knowledge Bank

About etown.com

CE Links

NEW EPISODE!

What is GPS?

Interactive Knowledge Bank

FLASH

Ida: Your Interactive Decision Assistant

**Initial recommendations**

Here's my initial suggested shortlist of the products that best meet your needs, based on what you've told me so far. They all offer an LCD view screen, manual overrides, an optical viewfinder, a serial output connection, a video out connection, a built-in digital zoom, and are within your requested price range.

- Olympus D502** See etown.com Reviews **\$499**  
Pros: It can store 18 pictures at its highest resolution, it has 1280 x 960 pixels resolution, it uses SmartMedia to store pictures, and it has an optical zoom lens.
- Olympus D400 Zoom** See etown.com Reviews **\$499**  
Pros: It can store 18 pictures at its highest resolution, it has 1280 x 960 pixels resolution, it uses SmartMedia to store pictures, and it has an optical zoom lens.
- Fujifilm MX1000** See etown.com Reviews **\$299**  
Pros: It can store 23 pictures at its highest resolution, it has 1280 x 960 pixels resolution, and it uses SmartMedia to store pictures. Cons: It doesn't have an optical zoom lens.
- Nikon D60** See etown.com Reviews **\$599**  
Pros: It has 1000 x 1200 pixels resolution, it has CompactFlash storage media, and it has an optical zoom lens. Cons: It can store only 8 pictures at its highest resolution.

I can refine these recommendations if you tell me more about your needs. I suggest **Optical zoom** as the next question to consider, or you can select the topic you wish.

Next question | Optical zoom

I can also show you a feature-by-feature comparison of any of the digital cameras listed on this page.

Compare Products

Other suitable products

Search

Products

Search

Advanced Search

QuickJump

Need expert advice? Ask Ida!

LIVE HELP

Click Here

to chat with our customer service

Customer Service

24 hours a day

7 days a week

Toll Free

877-GO-ETOWN

CHAT! Click Here

Audio Talk

Played LOUD!

Michael Bremer

Open 1 weekdays

The screenshot shows the etown.com website in a Microsoft Internet Explorer browser window. The page title is "etown.com - Ida Questions". The URL is "http://www.etown.com/sale\_bot/Ida.htm?pageType=secondary&SecondaryQuestion=31&question=17&question.y=11". The page features a navigation bar with "Customer Service | Your Account | Shopping Cart" and the date "July 17, 2000". The main heading is "Change the way you think...". On the left, there are "Categories" (Home Theater, Camcorders, Home Audio, Television, Portable Tech, Telecom) and "News / Views" (This Just In, Reviews/Products, Features/Columns, Community Hall, Knowledge Bank, About etown.com, CE Links). The central content area is titled "Ida: Your Interactive Decision Assistant" and focuses on "Optical zoom". It includes a cartoon character Ida and text explaining the difference between optical and digital zoom. A poll asks "How important is a zoom?" with three options: "Not very", "A digital zoom is fine", and "I need an optical zoom". The right sidebar contains a "Search" box, a "QuickJump" dropdown, a "Need expert advice? Ask Ida!" button, a "LIVE HELP" section with a "CLICK HERE" link, and a "Customer Service" section with "24 hours a day 7 days a week" and "Toll Free 877-GO-ETOWN". At the bottom, there is a "CHAT!" section with "Audio Talk" and "Play Loud!" buttons.

The screenshot shows the etown.com website in a Microsoft Internet Explorer browser window. The page title is "etown.com - Ida Questions". The URL is "http://www.etown.com/sale\_bot/Ida.htm?SecondaryQuestion=31&question=17&question.y=11". The page features a navigation bar with "Customer Service | Your Account | Shopping Cart" and the date "July 17, 2000". The main heading is "etown.com consumer electronics Summer Sale". On the left, there are "Categories" (Home Theater, Camcorders, Home Audio, Television, Portable Tech, Telecom) and "News / Views" (This Just In, Reviews/Products, Features/Columns, Community Hall, Knowledge Bank, About etown.com, CE Links). The central content area is titled "Ida: Your Interactive Decision Assistant" and focuses on "Recommendations". It includes a cartoon character Ida and text explaining the recommendations. A list of recommended products is shown, including "Olympus D450Z", "Olympus D460 Zoom", "FujifinePix1600", and "Nikon 880". The right sidebar contains a "Search" box, a "QuickJump" dropdown, a "Need expert advice? Ask Ida!" button, a "LIVE HELP" section with a "CLICK HERE" link, and a "Customer Service" section with "24 hours a day 7 days a week" and "Toll Free 877-GO-ETOWN". At the bottom, there is a "CHAT!" section with "Audio Talk" and "Play Loud!" buttons.

etown.com - Ida Questions - Microsoft Internet Explorer

Address: http://www.etown.com/index\_bat/ida.htm?pageType=explain&productID=3774

Customer Service | Your Account | Shopping Cart | July 17, 2000

**etown.com™** **NETFLIX DVD movie rentals**

**Categories**

- Home Theater
- Camcorders
- Home Audio
- Television
- Portable Tech
- Telecom

**News / Views**

This Just In  
[Beverly Products](#)  
[Features/Columns](#)  
[Community Hall](#)  
[Knowledge Bank](#)  
[About etown.com](#)  
[CE Links](#)

**NEW EPISODE!**  
 What is GPS?

**Interactive Knowledge Bank**  
[GET FLASH](#)

**Ida: Your Interactive Decision Assistant**

**Nikon 800 - \$599.00 (msrp)**  
 Given the information you have provided, the Nikon 800 is one of my top recommendations. Click the link to go directly to a question that will explain the feature and help you decide if that feature makes sense for you!

**Pros: Its advantages include:**

- it has an [LCD view screen](#).
- it has manual overrides.
- it has an [optical viewfinder](#).
- it has 1600 x 1200 pixels [resolution](#).
- it has a [serial output](#) connection.
- it has [CompactFlash](#) storage media.
- it has a [video out connection](#).
- it has a built-in digital zoom.
- it has an [optical zoom](#) lens.

**Cons: Possible disadvantages include:**

- it can [store](#) only 8 pictures at its highest resolution.
- it doesn't have a [USB connection](#).

\*The Nikon 800 is available for online purchase

[Return to Recommendations](#)

[Show full specification](#)

[Next Question](#)

**Search**

Products

Advanced Search

QuickJump

**Need expert advice?**  
**Ask Ida!**

**LIVE HELP**  
[CLICK HERE](#)  
 to chat with our customer service

**Customer Service**  
 24 hours a day  
 7 days a week

**Toll Free**  
 877 GO ETOWN

**CHAT!**

Audio Talk  
 Played LOUD  
 Michael Bremer  
 Sign - Mondays




Product Comparison - Microsoft Internet Explorer

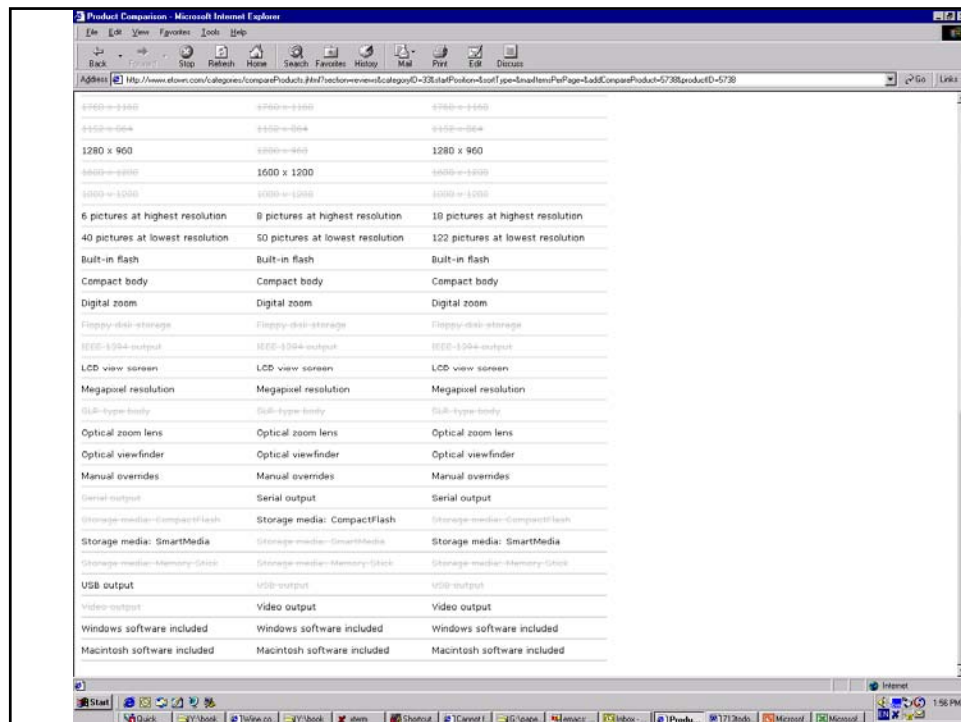
Address: http://www.etown.com/categories/compareProducts.htm?productReviewCategoryId=33&stateProductSortType=smallestPerPage=4&addCompareProduct=5738&productID=5738

**Your Product Comparisons**

Products are compared by key features. When a product lacks a feature, the feature appears in pale gray. To view a full Product Profile, click on the corresponding model number. To compare different products, click on Modify Comparison List (above).

**Digital Cameras:**

Fuji	Nikon	Olympus
		
<b>Model No.: FinePix 1900</b>	<b>Model No.: 800</b>	<b>Model No.: D450Z</b>
1.2-megapixel digital camera	CoolPix 2.1-megapixel digital camera	1.2-megapixel digital camera
1 year parts & labor	1 year parts & labor	1 year parts & labor
List Price: \$399.00	List Price: \$599.00	List Price: \$499.00
Low Price: \$329.00	Low Price: \$499.99	---
High Price: \$329.00	High Price: \$498.00	---
<a href="#">Buy Now!</a>	<a href="#">Buy Now!</a>	<a href="#">Find Local Retailer</a>
<a href="#">Find Local Retailer</a>	<a href="#">Find Local Retailer</a>	
320 → 240	320 → 240	320 → 240
640 → 480	640 → 480	640 → 480
690 → 500	690 → 500	690 → 500
1024 → 768	1024 → 768	1024 → 768
1200 → 1024	1200 → 1024	1200 → 1024
1536 → 1024	1536 → 1024	1536 → 1024
1544 → 1020	1544 → 1020	1544 → 1020



## Knowledge-Based Recommender

- Case-Based Example with Navigation Interface
- FindMe Systems (e.g., Entrée)

**Entree Results**

**The Los Angeles restaurant you chose is:**

**Chinois On Main**  
 2709 Main St. (bet Rose Ave. & Ocean Park Blvd.), Santa Monica, 310-392-9025  
 Pacific New Wave \$30-\$50  
 Extraordinary Decor, Extraordinary Service, Near-perfect Food, Hip Place To Be, On the Beach, Great for People Watching, Parties and Occasions, Weekend Brunch, Weekend Lunch, Fabulous Wine Lists

**We recommend:**

**Yoshi's Cafe**  
 3257 N. Halsted St. (Belmont Ave.), Chicago, 312-248-6160  
 Asian, Japanese, French (New) \$30-\$50  
 Extraordinary Decor, Extraordinary Service, Near-perfect Food, Need To Dress, Prix Fixe Menus, Quiet for Conversation, Very Busy - Reservations a Must, Romantic, Good Out of Town Business, Fabulous Wine Lists, Game, Parking/Valet

Less \$\$ nicer easier  
 traditional creative fancier quieter

**Entree Results**

**For a cheaper restaurant than:**

**Yoshi's Cafe**  
 3257 N. Halsted St. (Belmont Ave.), Chicago, 312-248-6160  
 Asian, Japanese, French (New) \$30-\$50  
 Good Decor, Excellent Service, Excellent Food, Creative, No Reservations, Weekend Brunch, Wheelchair Access, Long Drive

**We recommend:**

**Lulu's (map)**  
 626 Davis St. (bet Chicago & Orrington Aves.), Evanston, 708-869-4343  
 Japanese, Asian below \$15  
 Good Decor, Excellent Service, Excellent Food, Creative, No Reservations, Weekend Brunch, Wheelchair Access, Long Drive

Less \$\$ nicer easier  
 traditional creative fancier quieter

Figure 3: Navigation using the "Less \$\$" tweak

## More Generally

- Case-Based Approaches (Knowledge, Database, etc.) are often most helpful for ephemerally-personalized experiences
  - Shopping – suggest similar relevant items
    - Compare with collaborative – suggest items that are co-purchased or co-browsed
  - Content – suggest similar stories
- Case-Based recommendations are often easier to explain to the user

## Challenges and Drawbacks

- Content-Based Techniques in general ...
  - Depend on well-structured attributes that align with preferences (consider paintings)
  - Depend on having a reasonable distribution of attributes across items (and vice versa)
  - Unlikely to find surprising connections (e.g., chili peppers or lemon with chocolate)
  - Harder to find complements than substitutes

## Some take-away lessons

- Many ways to recommend based on content (product attributes)
  - Long-term: build profile of content preferences
  - Shorter-term: build database of cases; navigate
- Content-based techniques work without a large set of users (but need item data)
- Good at finding substitutes; good at helping navigate for a purchase; good explainability

## Moving Forward

- Lectures
  - Details of TFIDF and Content-Based Filtering
  - Extensions
  - Interviews with experts on case-based, knowledge-based
  - Evolution from content-based to vector-space approaches
  - Programming CBF in LensKit (Honors Track)



# Introduction to Content-Based Recommenders