Exploiting and Learning Temperament-Based Information Filtering

利用與學習基於個性的資訊過濾法

林葭華

國立中山大學資訊工程學系

chlin@cse.nsysu.edu.tw

民國**92**年**10**月**1**日

Outline

- Motivation & Goal
- Temperament-Based Filtering Method
  - Temperament
  - Segmentation
  - Architecture
  - Learning
  - Classification
  - Filtering
  - Relevance Feedback
- Experimental Implementations
- Future Work & Conclusions

Motivation

- **Rapid proliferation of on-line digital information**
  - **New information customization techniques are needed to assist people**
- **Limitations of existing information filtering techniques**
  - **Inability to satisfy the user**
  - **E.g., Content-based filtering and Social filtering**
- **Challenging issues in clustering techniques**
  - **Identify concepts in a knowledge base**
  - **Construct internal representations for partitioning the diverse concepts into appropriate categories**

Content-Based Filtering

- **Keyword-based document/information retrieval systems**
  - **Draw on vector space model**
  - **E.g., Web search engines**
- **Shortfalls**
  - **Information units cannot be fully featured by key terms**
    - E.g., image, audio, video, art, or physical items
  - **Items cannot be filtered on quality, style, or point-of-view**
  - **Not suitable for serendipitous search**
    - Require the user to know what they like exactly

Vector Space Model

Document vectors

$$
\begin{array}{c} \\ \texttt{Doc } 1 \\ \texttt{Doc } 2 \\ .. \\ \texttt{Doc } n \end{array}
\begin{array}{c}
\begin{array}{cccc} \texttt{Term } 1 & \texttt{Term } 2 & .. & \texttt{Term } m \end{array} \\
\begin{pmatrix}
x_{11} & x_{12} & .. & x_{1m} \\
x_{21} & x_{22} & .. & x_{2m} \\
.. & .. & .. & .. \\
x_{n1} & x_{n2} & .. & x_{nm}
\end{pmatrix}
\end{array}
$$

- **User requests and document contents are represented by term vectors**
  - **Cell x_{ij} is the weight of term j assigned to vector i**
- **TF- IDF (Term Frequency and Inverse Document Frequency) weighting method**
- **Cosine similarity measure**
  - **Calculate the similarity btw any two vectors**
  - **Group vectors into clusters**

TF-IDF (Term Frequency and Inverse Document Frequency) Method

A term weight is defined as

$$W_i = \frac{TF_i \times IDF_i}{\sqrt{\sum_{j=1}^{n} TF_j^2 \times IDF_j^2}} \quad \textbf{and} \quad IDF_i = \log_2\left(\frac{n}{DF_i}\right)$$

where

TF_i = **the number of times term or concept m_i appears in information unit d (the term frequency)**

DF_i = **the number of information units in the collection which contain m_i (the document frequency)**

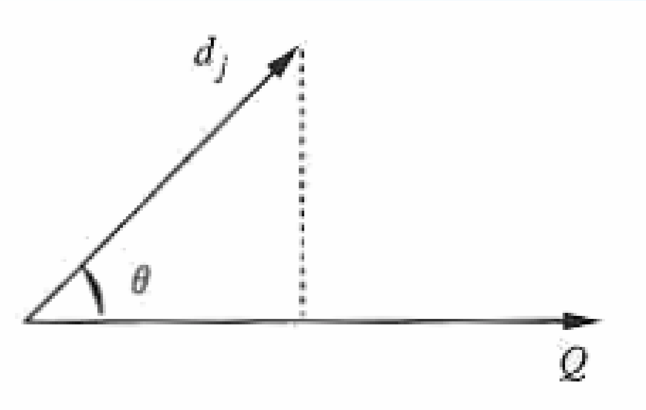
IDF_i = **the inverse document frequency**

n = **the number of information units in the collection**

Cosine Similarity Measure

- **Term weights are used to compute the degree of similarity between documents and the user query**

$$Sim(d_j, Q) = \frac{\vec{d}_j \cdot \vec{q}}{|\vec{d}_j| \times |\vec{q}|} = \frac{\sum_{i=1}^{t} w_{i,j} \times w_{i,q}}{\sqrt{\sum_{i=1}^{t} w_{i,j}^2} \times \sqrt{\sum_{i=1}^{t} w_{i,q}^2}}$$



$$\vec{d}_j = (w_{1,j}, w_{2,j}, \cdots, w_{t,j})$$
$$\vec{q} = (w_{1,q}, w_{2,q}, \cdots, w_{t,q})$$

- **The cosine of θ is adopted as *Sim(d$_j$,Q)***

Social Filtering

- **Recommend information by evaluations or opinion of other users**
  - **Based on similarities between the user interest (rating) profiles by statistical analysis**
- **Shortfalls**
  - **No predefined concept classes are used to describe or simplify the meaning of the clusters**
  - **No content analysis is used to represent the textual features of an information unit**

Hypothesis & Goal

- **The accuracy of an information recommendation system can be significantly improved by employing user temperament for filtering and customization**
- **To characterize the information space by taking human factors into consideration and devise a new filtering mechanism to provide a better information recommendation service**

Human Temperament

- A predominant factor in the patterns of human behaviors and preferences
  - Psychologists
- An innate property of the brain
  - Neuroscience research
- Relevance to the public taste
  - Some of the studies

    In perceiving or interpreting the information in general
  - “Men of each psychological type tend to admire the art produced by artists of the same type” [Evans 1939]
- Strong potential for an effective information filtering

Research Approach

- **New filtering mechanism**
  - **Temperament-based**
  - **Intelligent multiagent collaborative system**
  - **Address segmentation, learning, classification, and filtering techniques**
  - **Heuristic selection rules**
- **Basic theories**
  - **Keirsey's temperament theory**
  - **Probability theory**
    - Distributions of temperaments
  - **Statistical reasoning**

Outline

- Motivation & Goal
- *Temperament-Based Filtering Method*
  - Temperament
  - Segmentation
  - Architecture
  - Learning
  - Classification
  - Filtering
  - Relevance Feedback
- Experimental Implementations
- Future Work & Conclusions

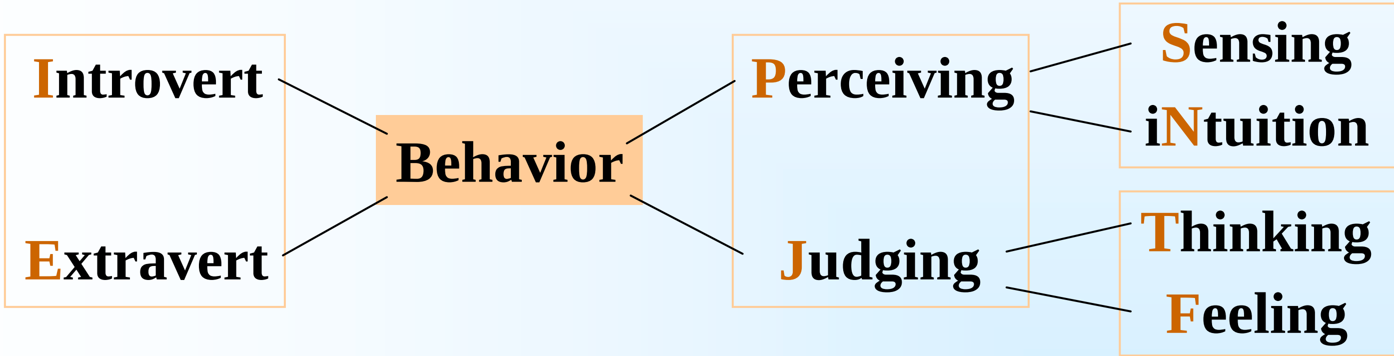
Temperament Theory

- **Temperament and Psychological Type**
  - **Represent two systems of classifying personalities, which later converged**
- **Hippocrates first named four temperaments**
  - **Choleric, phlegmatic, melancholic, and sanguine**
- **David Keirsey derived new theory**
  - **Hippocrates' four temperaments are misleading**
  - **Correlated into the Myers-Briggs Type system**
  - **Classified the sixteen personality types into four temperaments as SJ, SP, NT, and NF**

Psychological Type Theory

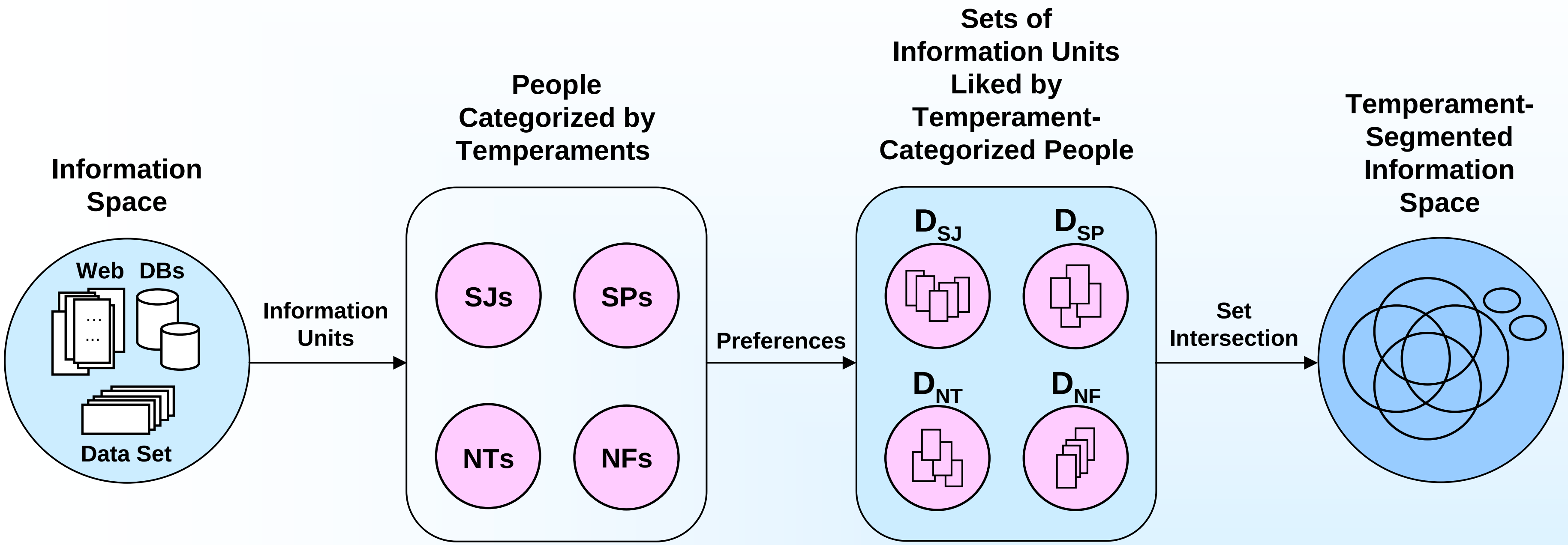
- **Carl Jung asserted people are fundamentally different**
  - **Classified into "psychological types"**
  - **Characterized by four pairs of opposite preferences**

    Within each pair, a person leans toward one or the other



- **Katherine Briggs & Isabel Myers adopted Jung's theory**
  - **Designed the Myers-Briggs Type Indicator (MBTI)**
  - **A type is represented by a four-letter code of personal preferences (e.g., ESTJ)**

Temperament-Segmented Feature of the Real-World Information Space



- **The percentage distributions of the temperaments in the United States** [Hammer 96]

Temperament (%)	MBTI (%)			
SJ 46.7	ESTJ 9.9	ESFJ 9.6	ISTJ 15.6	ISFJ 11.5
SP 21.4	ESTP 4.8	ESFP 5.7	ISTP 6.4	ISFP 4.5
NT 16.1	ENTJ 2.8	ENTP 4.7	INTJ 3.5	INTP 5.2
NF 15.8	ENFJ 2.5	ENFP 6.3	INFJ 2.6	INFP 4.3

- **Segmentation Function**

$$S_n = \bigcap_{t \in T_n} D_t - \sum_{t' \in (T - T_n)} D_{t'}$$

$$S_1 = D_{SJ} \cap D_{SP} \cap D_{NT} \cap D_{NF}$$

$$S_2 = D_{SJ} \cap D_{SP} \cap D_{NT} - D_{NF}$$

$$\vdots$$

$$S_{16} = D_{\phi}$$

Segmentation Function

The segment of an information space may be defined as

$$S_n = \bigcap_{t \in T_n} D_t - \sum_{t' \in (T - T_n)} D_{t'}$$

where

n **=** $\{1, 2, \ldots, 16\} = \{1, 2, \ldots, 2^{|T|}\}$

T **=** $\{SJ, SP, NT, NF\}$, $|T|$ **= size of** T

T_n **=** $\{t \mid t$ **is a temperament value in segment** $S_n\}$

D_t **= the set of information units evaluated as "like" by users with temperament** t

and

$D_{t'}$ **= the set of information units evaluated as "like" by users with temperament** t'

Sample Segments & the Relevant Temperaments

$$S_1 = D_{SJ} \cap D_{SP} \cap D_{NT} \cap D_{NF} \qquad T_1 = \{SJ, SP, NT, NF\}$$

$$S_2 = D_{SJ} \cap D_{SP} \cap D_{NT} - D_{NF} \qquad T_2 = \{SJ, SP, NT\}$$

$$S_3 = D_{SJ} \cap D_{SP} \cap D_{NF} - D_{NT} \qquad T_3 = \{SJ, SP, NF\}$$

.

.

.

$$S_{14} = D_{NT} - D_{SJ} - D_{SP} - D_{NF} \qquad T_{14} = \{NT\}$$

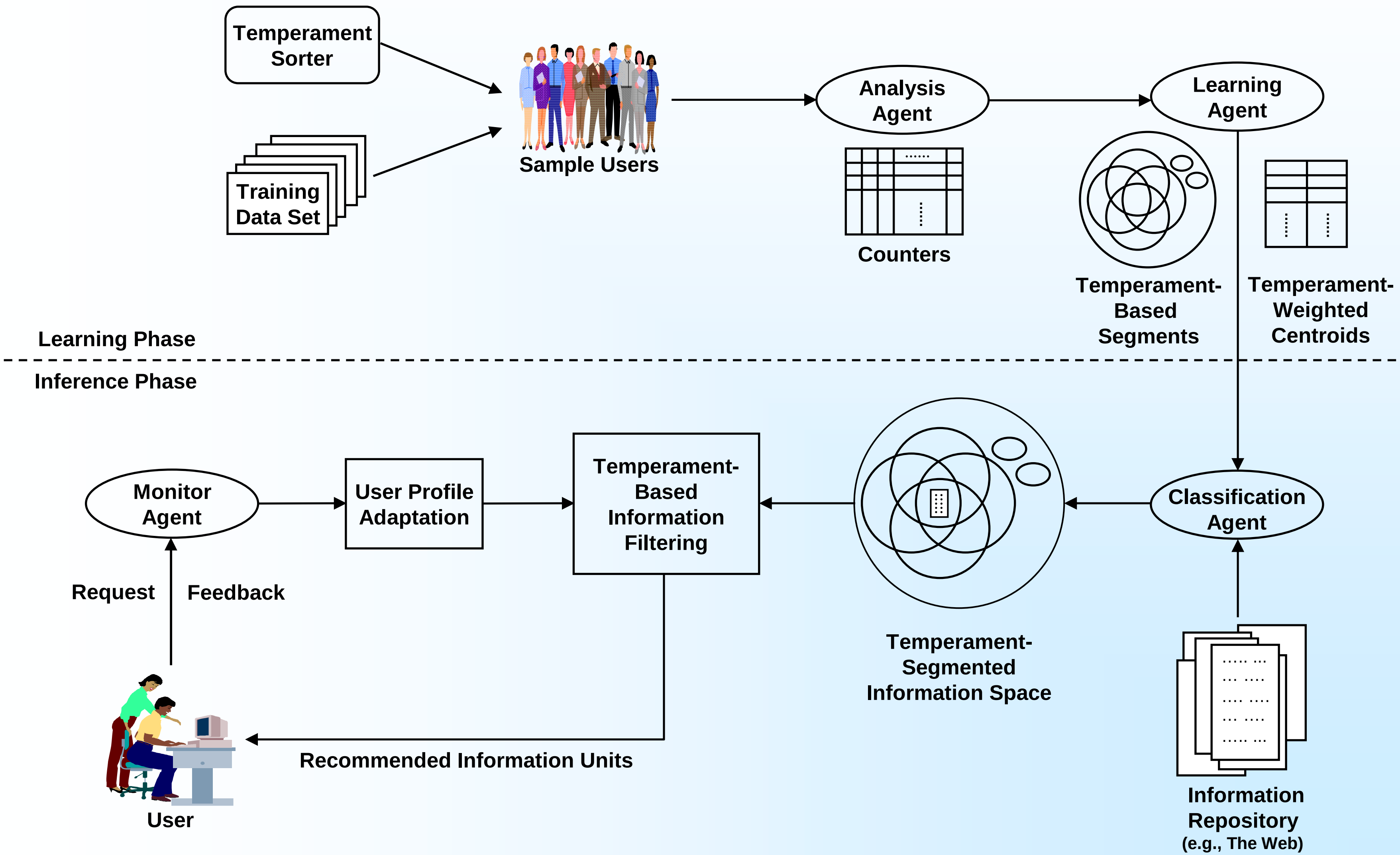
$$S_{15} = D_{NF} - D_{SJ} - D_{SP} - D_{NT} \qquad T_{15} = \{NF\}$$

$$S_{16} = D_{\phi} \qquad T_{16} = \phi$$

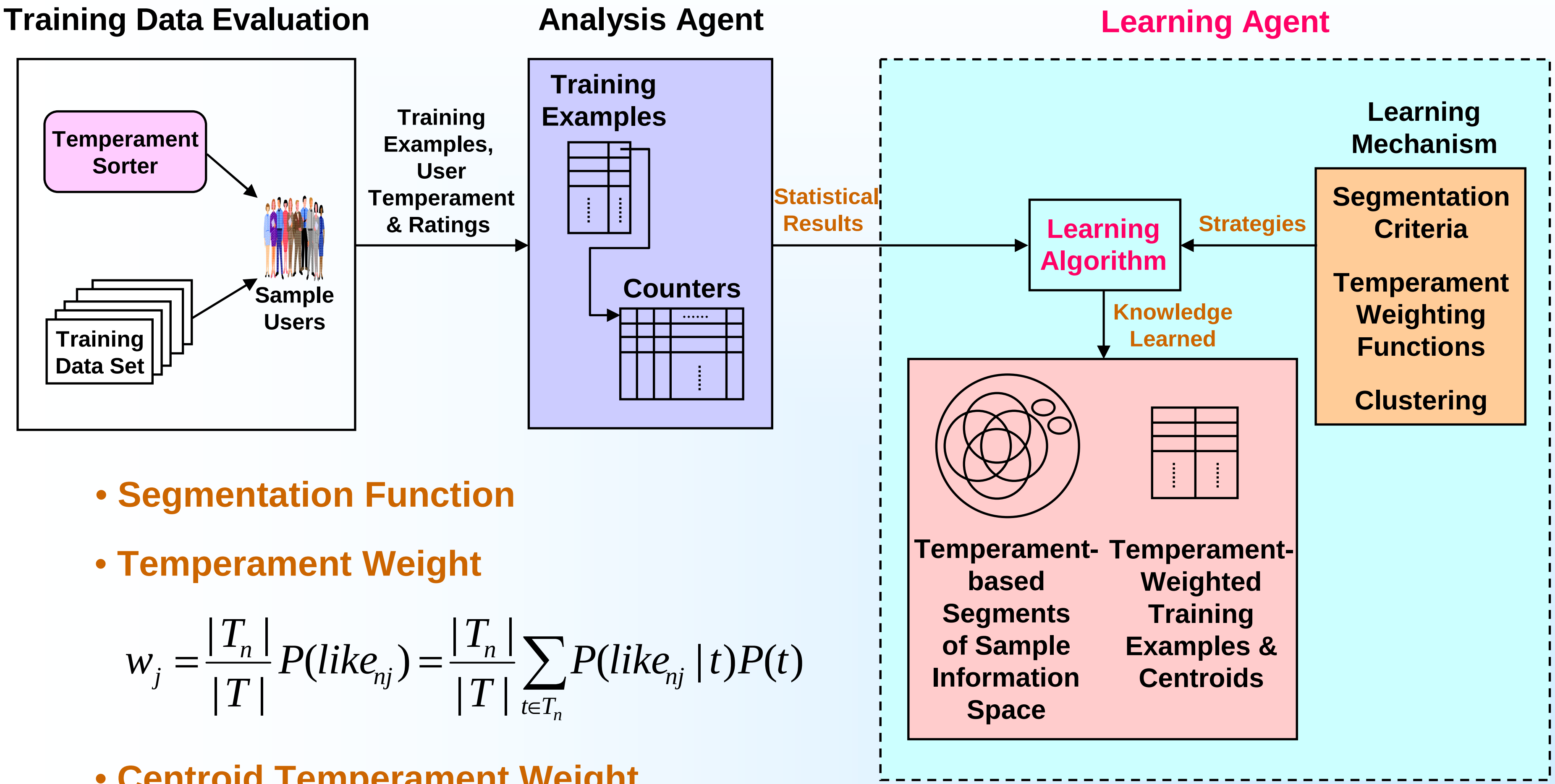
where

$D_{\phi} =$ **the set of information units not evaluated as "like" by any user**

Architecture of the Temperament-Based Information Filtering Model



Learning Process



- Segmentation Function
- Temperament Weight

$$w_j = \frac{|T_n|}{|T|} P(like_{nj}) = \frac{|T_n|}{|T|} \sum_{t \in T_n} P(like_{nj} \mid t) P(t)$$

- Centroid Temperament Weight

$$e_i = \frac{1}{m} \sum_{j=1}^{m} w_j$$

Segmentation Criteria

- **Segmentation function** $S_n = \bigcap_{t \in T_n} D_t - \sum_{t' \in (T - T_n)} D_{t'}$
- **Eliminate the biased data or false alarm**
  - **An information unit is learned to be in one of the segments except S_{16} only when sufficient evidences have been observed to pass a preset threshold**

    Otherwise, the item is retained in S_{16}
- **Systematic search**
  - **Segments are sorted in descending order of their accumulated probability of temperaments,** $\sum_{t \in T_n} P(t)$
  - **A segment with a lower index will contain information units having a higher accumulated probability or popularity**

Segments of a Temperament-Based Sample Information Space

- **Popularity Threshold**

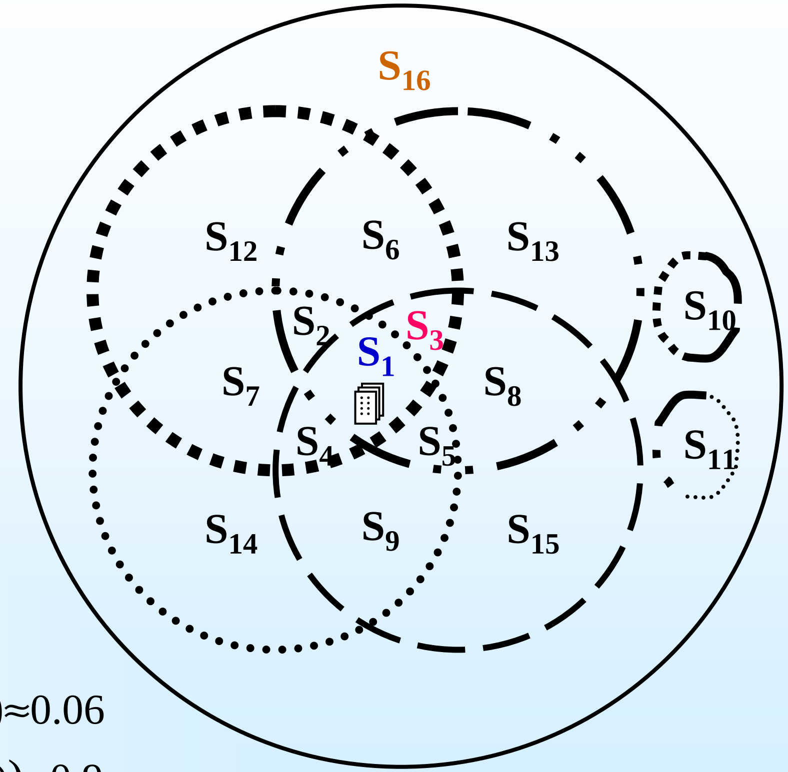
  - Precondition to form the segments

$$P(like_{dj} \mid t) > \theta$$

- **Segmentation Function**

$$S_n = \bigcap_{t \in T_n} D_t - \sum_{t' \in (T - T_n)} D_{t'}$$

$$S_1 = D_{SJ} \cap D_{SP} \cap D_{NT} \cap D_{NF}$$
$$S_2 = D_{SJ} \cap D_{SP} \cap D_{NT} - D_{NF}$$
$$S_3 = D_{SJ} \cap D_{SP} \cap D_{NF} - D_{NT}$$
$$\vdots$$
$$S_{16} = D_{\phi}$$



Legend:
Information Space ———
D_{SJ} ▪▪▪▪▪▪ D_{SP} — · · —
D_{NT} ·········· D_{NF} — — —

Suppose $\theta = 0.1$, then we have ($d1$, $S3$)

$P(like_{d1}|SJ)=(19/(19+1))\approx 0.95$ $P(like_{d1}|NT)=(1/(1+15))\approx 0.06$

$P(like_{d1}|SP)=(10/(10+2))\approx 0.83$ $P(like_{d1}|NF)=(18/(18+2))\approx 0.9$

Information Unit	Segment #	SJ		SP		NT		NF	
		Yes	No	Yes	No	Yes	No	Yes	No
d1	S3	19	1	10	2	1	15	18	2
d3	S16	0	20	0	5	0	5	0	10
d6	S1	50	0	30	0	10	0	10	0

(Scale of # of User Responses: 100)

Temperament Weight

- **Purpose**
  - **To quantify and estimate the relative influence of various temperaments on the popularity measure of an information unit**
- **Basic ideas**
  - **An Information unit with a higher prior probability indicates the likelihood of a larger interested population**
  - **A segment consisting more temperament types exhibits more human diversities in the interested population**

    E.g., segments have information units with the same prior probability

$\Rightarrow$ **May have heavier weight**

Estimate Temperament Weight

For an information unit d_j in segment S_n

$$w_j = \frac{|T_n|}{|T|} P(like_{nj}) = \frac{|T_n|}{|T|} \sum_{t \in T_n} P(like_{nj} \mid t) P(t)$$

where

T_n **= {** $t \mid t$ **is a temperament value in segment** S_n **}**

T **= {*SJ, SP, NT, NF*},** $|T_n|$ **= size of** T_n**,** $|T|$ **= size of** T

$P(like_{nj})$ **= the prior probability that an information unit** d_j **in segment** S_n **is evaluated as "like"**

$P(like_{nj}|t)$ **= the conditional probability that an information unit** d_j **in segment** S_n **is evaluated "like", given user temperament** t

$P(t)$ **= the percentage distribution of temperament** t

and $\sum_{t \in T_n} P(t) = 1$ **(temperaments are mutually exclusive)**

Estimate Temperament Weights of Information Units - An Example

Weighting Function

$$w_j = \frac{|T_n|}{|T|} \sum_{t \in T_n} P(like_{nj} | t) P(t)$$

w_1 = (3/4)(19/20*0.467+10/12*0.214+18/20*0.158)=0.5731

w_2 = (1/4)(10/20*0.161)=0.0201

w_3 = 0

w_4 = (2/4)(10/10*0.467+40/40*0.161)=0.314

w_5 = (4/4)(10/10*0.467+20/30*0.214+10/20*0.161+10/20*0.158)=0.7692

w_6 = (4/4)(50/50*0.467+30/30*0.214+10/10*0.161+10/10*0.158)=1

Information Unit	# of User Responses	SJ		SP		NT		NF	
		Yes	No	Yes	No	Yes	No	Yes	No
d1	62	19	1	10	2	0	10	18	2
d2	45	0	5	0	10	10	10	0	10
d3	40	0	20	0	5	0	5	0	10
d4	90	10	0	0	20	40	0	0	20
d5	80	10	0	20	10	10	10	10	10
d6	100	50	0	30	0	10	0	10	0

Statistical Results (Scale of # of User Responses: 100)

Centroid Temperament Weight

- **An estimate of the popularity within the user population for a cluster in a segment**
- **The centroid vector is defined as the mathematical average of the information unit vectors in the cluster**
- **The centroid temperament weight is defined as**

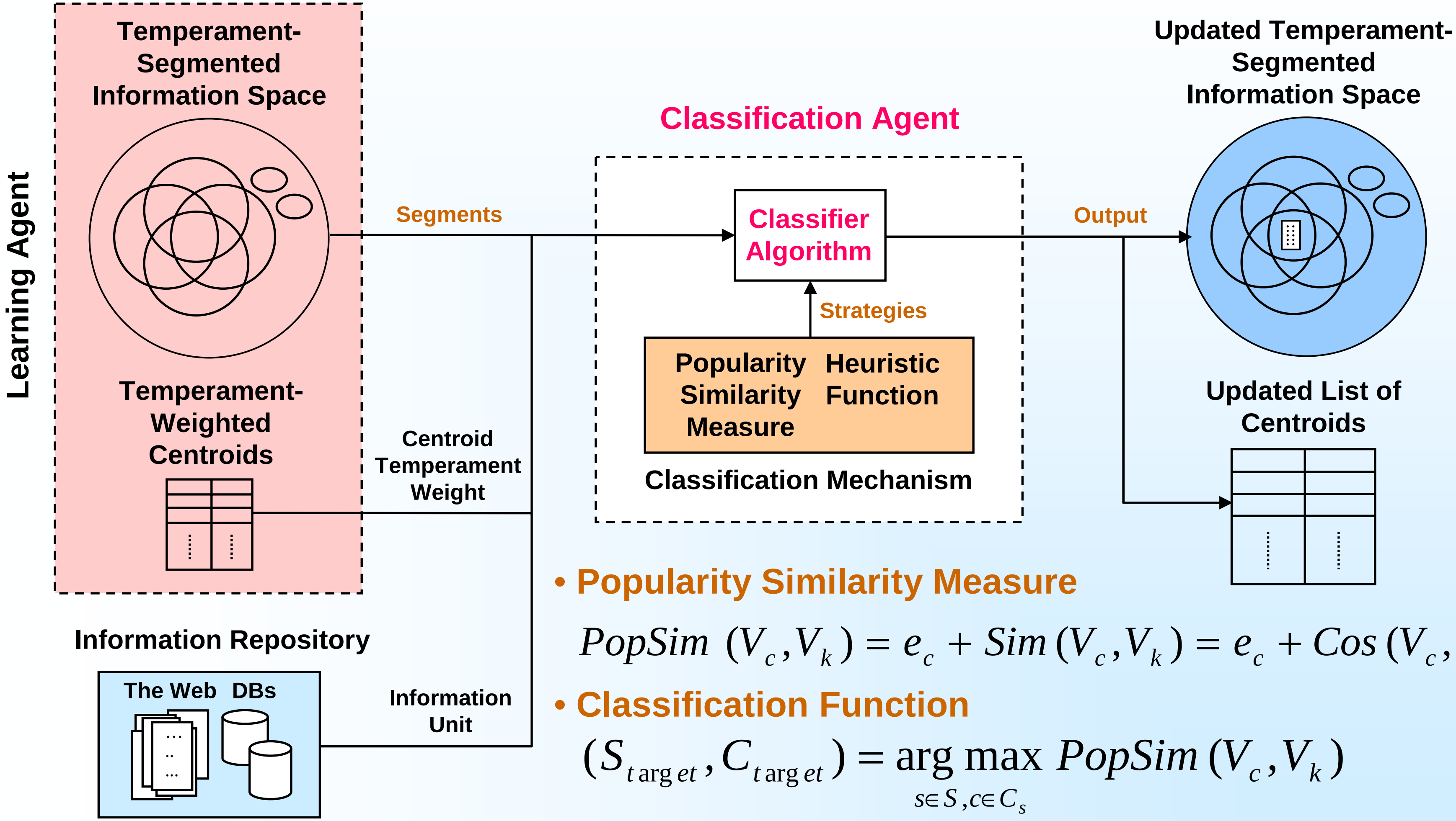
$$e_i = \frac{1}{m}\sum_{j=1}^{m} w_j$$

where

m = **size of cluster** c_i

w_j = **the temperament weight of an information unit** d_j **in cluster** c_i

Classification Process



- **Popularity Similarity Measure**

$$PopSim\ (V_c, V_k) = e_c + Sim\,(V_c, V_k) = e_c + Cos\,(V_c, V_k)$$

- **Classification Function**

$$(S_{target}, C_{target}) = \underset{s \in S,\, c \in C_s}{\arg\max}\ PopSim\,(V_c, V_k)$$

- **The location of a new information unit can be adapted by observing user feedback for that unit**

Temperament-Based Classification

- Assumption
  - A widely liked set of information units by a particular group of people
  - A new information unit having content features similar to that set of information units

  ⇒ The new information unit is probably liked by that group of people
- Basic idea
  - A higher centroid temperament weight indicates that the information units in that cluster have a greater popularity among the user population

  ⇒ Centroid temperament weight is an important factor for popularity measure

Popularity Similarity Measure

Estimate the level of importance of both popularity and similarity

Definition

$$PopSim(V_c, V_k) = e_c + Sim(V_c, V_k)$$

where

V_c = **the centroid vector of a temperament-classified cluster**

V_k = **a target vector (new information unit or user query)**

e_c = **the temperament weight of** V_c

$Sim(V_c, V_k)$ = **the traditional cosine similarity measure**

$$= \frac{V_c \cdot V_j}{|V_c| \times |V_j|}$$

The Classifier

Classification function

$$(S_{target}, C_{target}) = \underset{s \in S, c \in C_s}{\arg\max}\, PopSim(V_c, V_k)$$

where

S = {s | s **is a segment of a temperament-based information space**}

C_s = {c | c **is a cluster of segment** s}

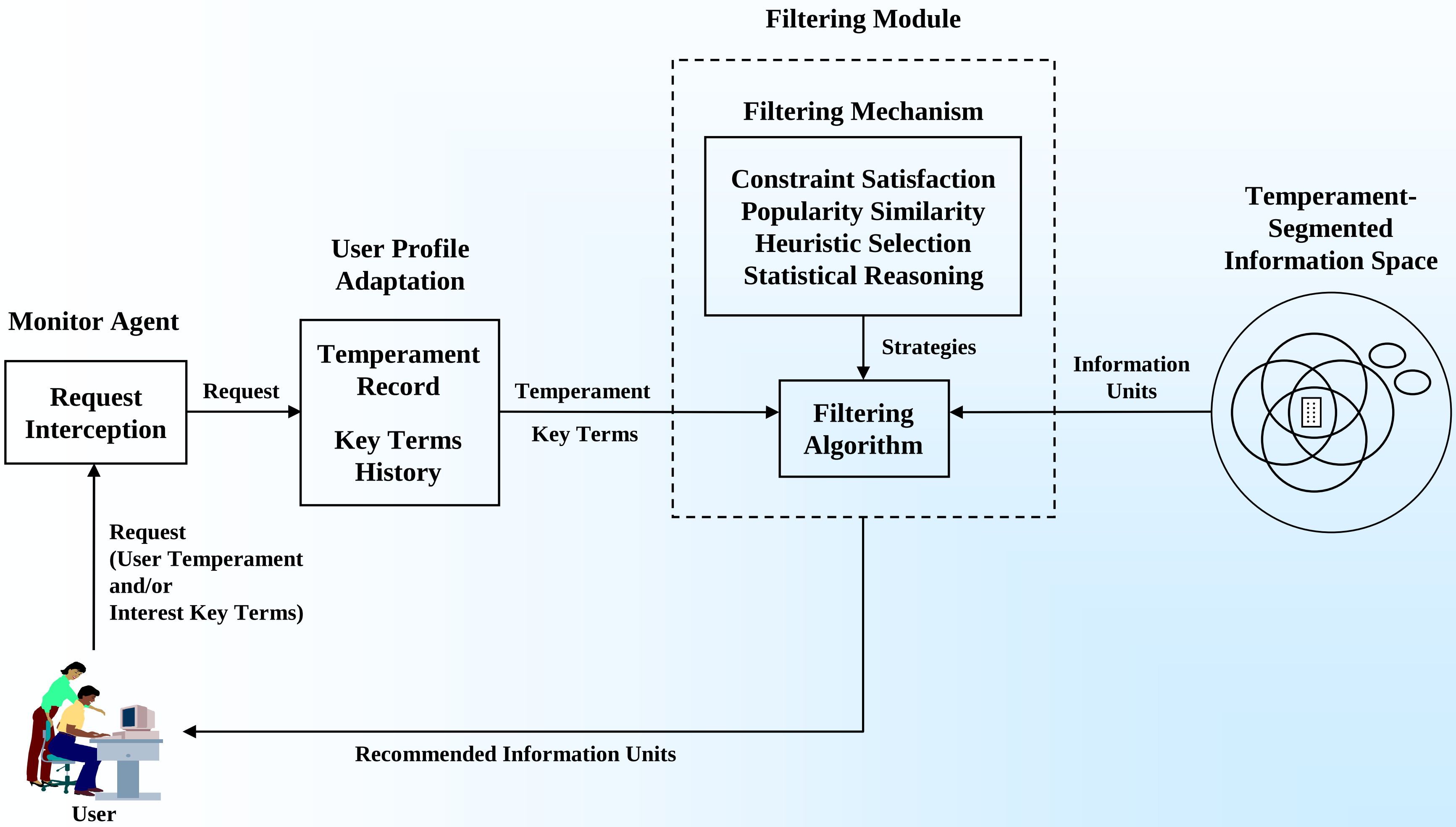
V_c = **the centroid vector of a temperament-classified cluster**

V_k = **a new information unit vector**

$PopSim(V_c, V_k)$ = **the popularity similarity of** V_c **and** V_k

- **The location of the new information unit can be adapted by observing user feedback for that unit**

Filtering Process



Temperament-Based Filtering

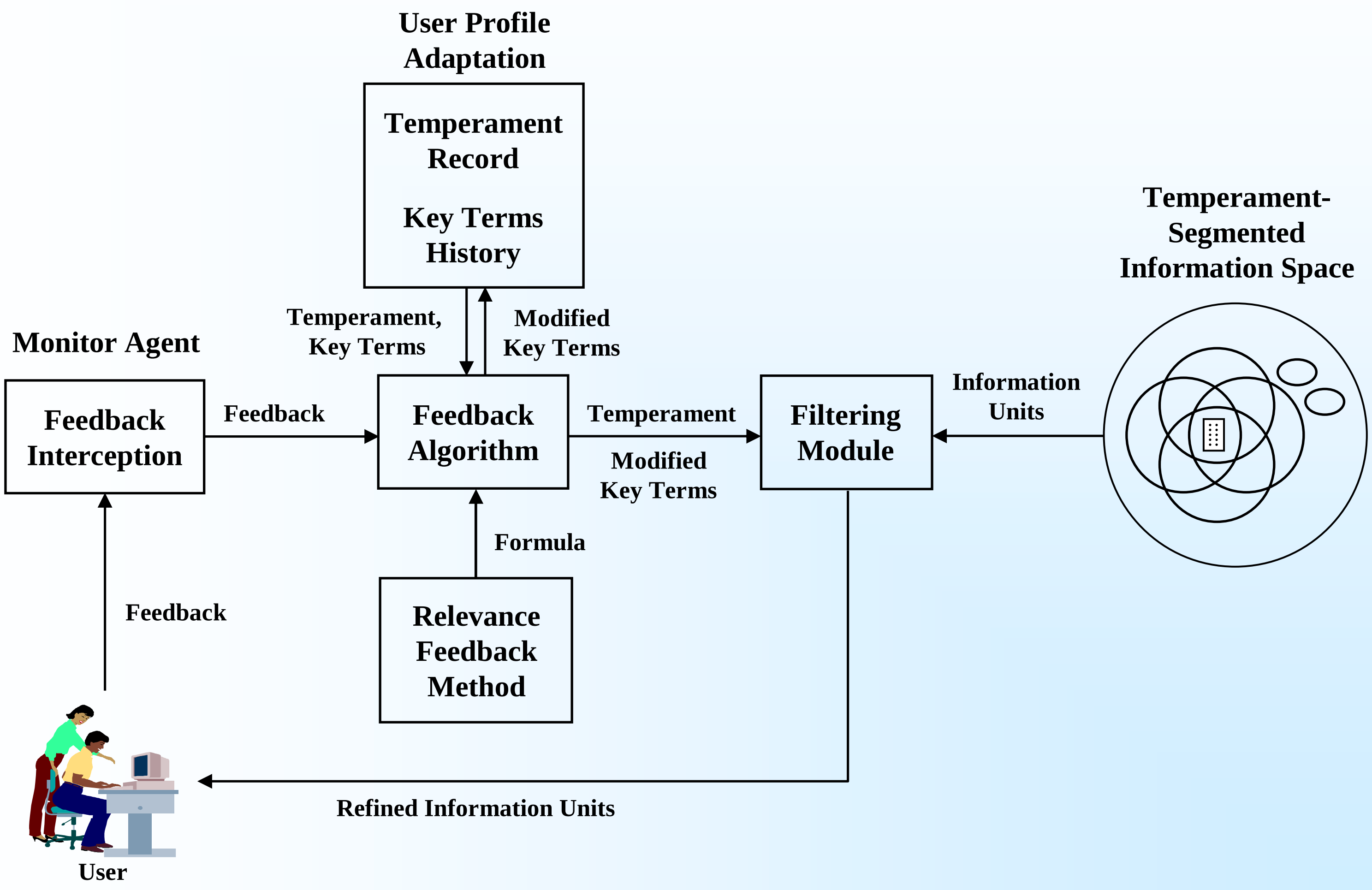
- **Assumption**
  - **A widely liked set of information units by a particular temperament type of people**
  - **A user having that temperament type**

  $\Rightarrow$ **The user would probably like the information units in that set**

Heuristics in Filtering Process

- **User with unknown temperament and interest**
  - **Return the items in segment S_1**
- **User temperament constraint**
  - **Search only the partial space**
- **User interest (key terms, V_k) constraint**
  - **Select items from the optimal target (segment, cluster)**
- **User temperament and interest constraint**
  - **Search only the partial space and select items from the optimal target (segment, cluster)**
- **Search only the upper (higher popularity) quartile segments in the information space at quick search mode**

Relevance Feedback Process



Relevance Feedback

- **Basic idea - Move toward "like" ~~and from "dislike"~~**
  - **Negative relevant feedback can be omitted [Aalbersberg 1992]**
- **Adopt and modify Ide Dec-Hi Method**
  - **Very efficient [Salton 1990]**
- **The modified interest (term keys) vector is defined as**

$$V_{k+1} = V_k + \sum_{V_r^k \in D_r^k} V_r^k \cancel{- V_{n1}^k}$$

where

V_k **= the original interest (term keys) vector**

D_r^k **= the set of relevant information units retrieved for interest vector**

V_r^k **= a relevant information unit in** D_r^k

Outline

- **Motivation & Goal**
- **Temperament-Based Filtering Method**
  - **Temperament**
  - **Segmentation**
  - **Architecture**
  - **Learning**
  - **Classification**
  - **Filtering**
  - **Relevance Feedback**
- ***Experimental Implementations***
- **Future Work & Conclusions**

Experimental Testing

- **Prototype System**

  To evaluate the proposed temperament-based filtering approach in comparison with content-based filtering method

- **Performance Metrics**

  An experimental user-studies testing procedure was conducted to demonstrate the improved effectiveness, by utilizing a combination of simulation and user-studies testing

- **Performance Evaluation**

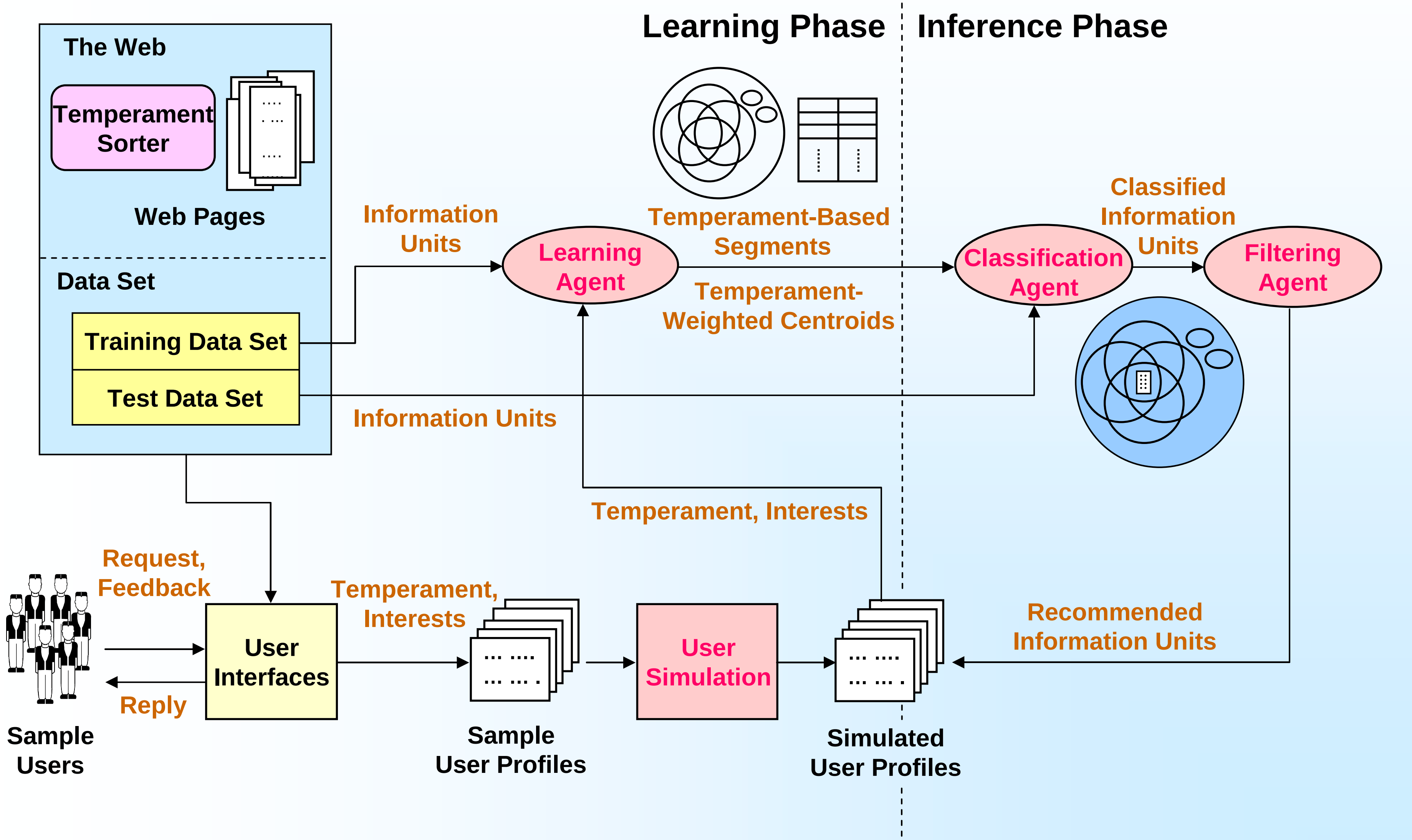
  The prototype information recommendation system was applied to and demonstrated in three application domains

  – A document collection

  – An art image collection

  – Representation styles

Architecture of the Temperament-Based Information Filtering Prototype System



Experimental Evaluation Measures

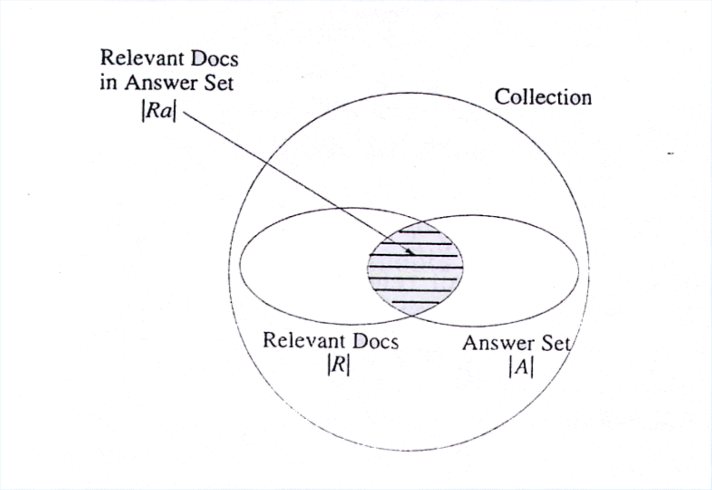
- **Percentage Accuracy**
  - **The fraction of positive user feedback in the top n items recommended by the system**
- **Normalized Distance-based Performance Measure**
  - **Distance between the user's ranking and the system's ranking of the same set of documents**
  - **Normalized to be between 0 and 1**
- **Precision and Recall**
  - **Precision is the proportion of retrieved material that is relevant**
  - **Recall is the proportion of relevant material retrieved**

Precision and Recall

- **Precision is the fraction of the retrieved docs** $|A|$ **which is revelant** $|Ra|$
- **Recall is the fraction of the relevant docs** $|R|$ **which has been retrieved** $|Ra|$

$$\text{Precision} = \frac{|Ra|}{|A|}$$

$$\text{Recall} = \frac{|Ra|}{|R|}$$



Experimental Evaluation Methodology

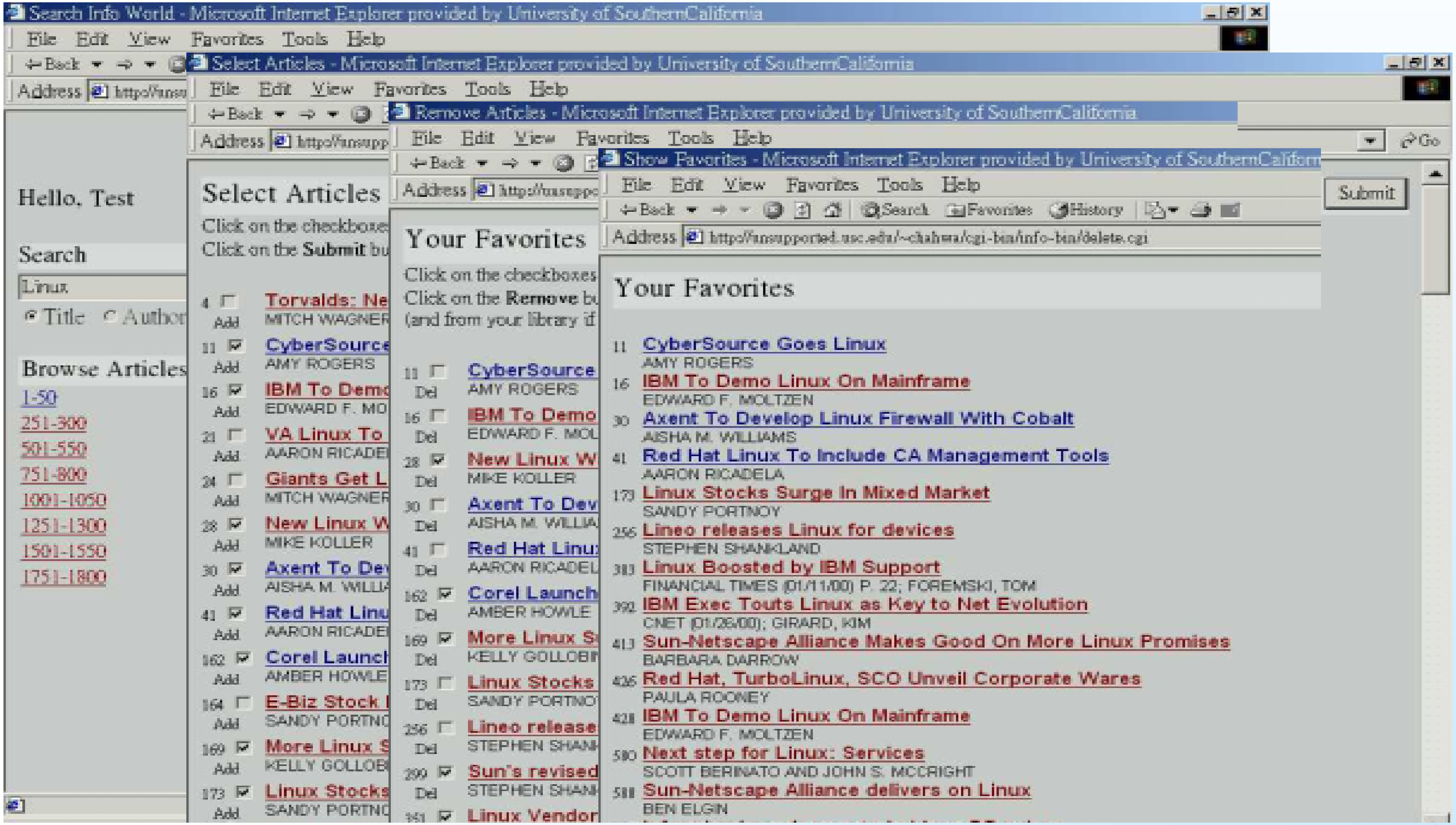
Training data set and User ratings

	← Test examples →			← Training examples →				
	d1		d100	d101	d102		d1999	d2000
User 1	..		..	..	..		..	..
User 2	..		..	..	..		..	..
.	..		..	..	..		..	..
.	..		..	..	..		..	..
User n	..		..	..	..		..	..

- **The training data set will be split into 20 sessions**
  - One session will be used as the test examples and the remaining 19 sessions as the training examples
  - Each session will be used as the test examples in one of the 20 experiments
  - The results of these 20 experiments will be averaged
- **The same training data set is used to evaluate the other filtering methods**

Experiment #1
A Document Collection of 2,000 Web Pages

- **http://www.usc.edu/dept/cs/tfiltering/infoworld**

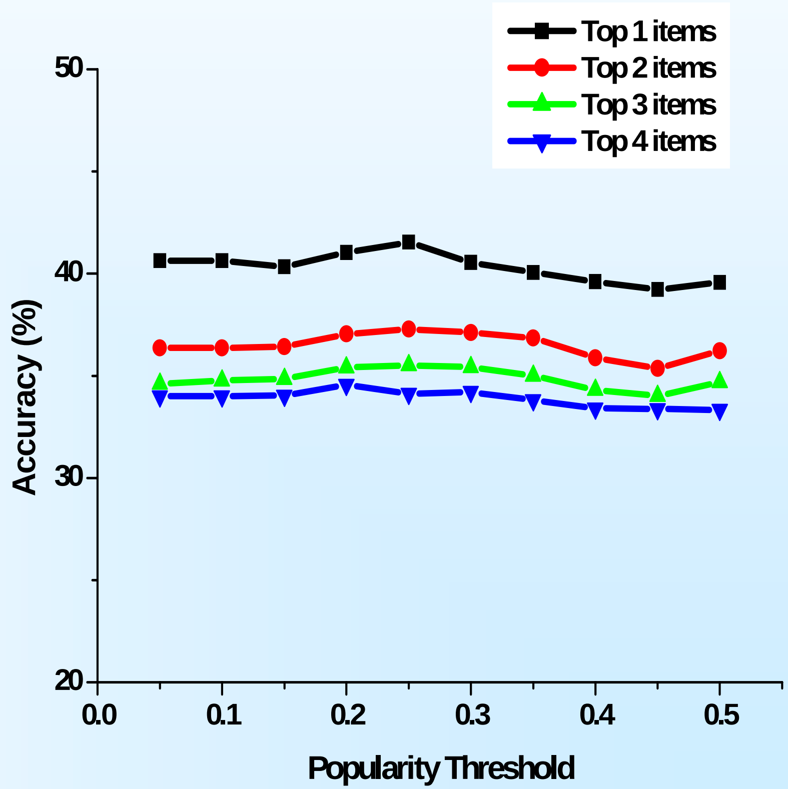


Fixing a Popularity Threshold for Segmentation in Temperament-Based Filtering

- **Examine popularity threshold θ = 0.05 to 0.55**
  - **Accuracy of Recommendation (20 tests)**
  - **Users with unknown temperament and interest**
- **Accuracy is stable**
  - **Top 4 items recommended**
- **Base threshold θ = 0.1**
  - **Exclude initial bias effect**
  - **Moderate size of a segment**

- **Popularity Threshold**

$$P(\, like_{dj} \mid t \,) > \theta$$



Fixing a Cosine Similarity Threshold for Clustering

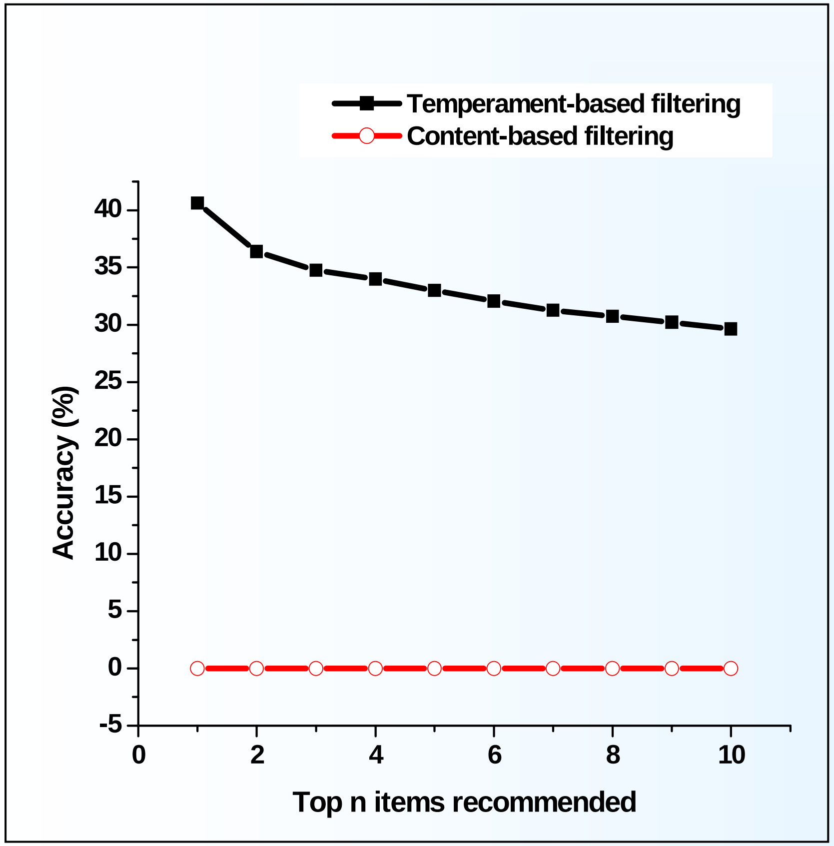
- **Information units are grouped into clusters**
  - **Cosine similarity measure > some threshold λ**
  - **Reduce the size of comparisons**
  - **Facilitate search**
- **About 10 units per cluster in content-based filtering**
  - **λ = 0.07, 0.15, and 0.04 for experiments #1, #2, and #3, respectively**
  - **Same λ for temperament-based filtering**
  - **No λ value for experiment #4 on Representation styles**
    - **Content-based filtering is not applicable**
    - **Exploit segments only**

Experiment #1 - Cases (a) & (b)

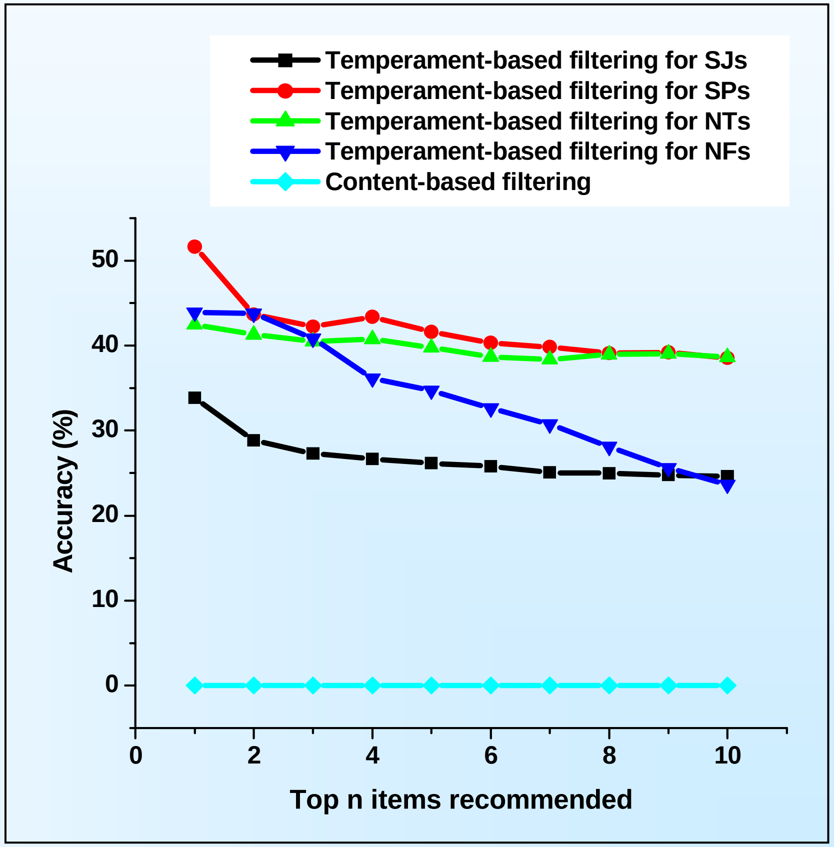
- **Accuracy of Recommendation (2000 documents, 20 tests)**

  (a) Users with unknown temperament and interest

  (b) Given the user temperament



(a)



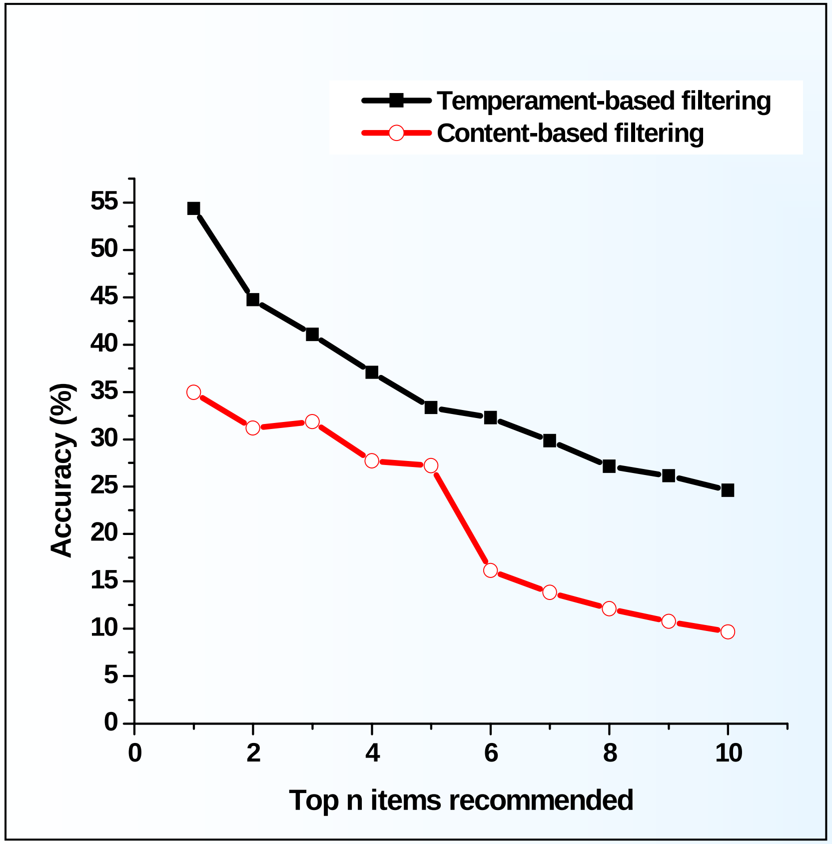
(b)

Experiment #1 - Cases (c) & (d)

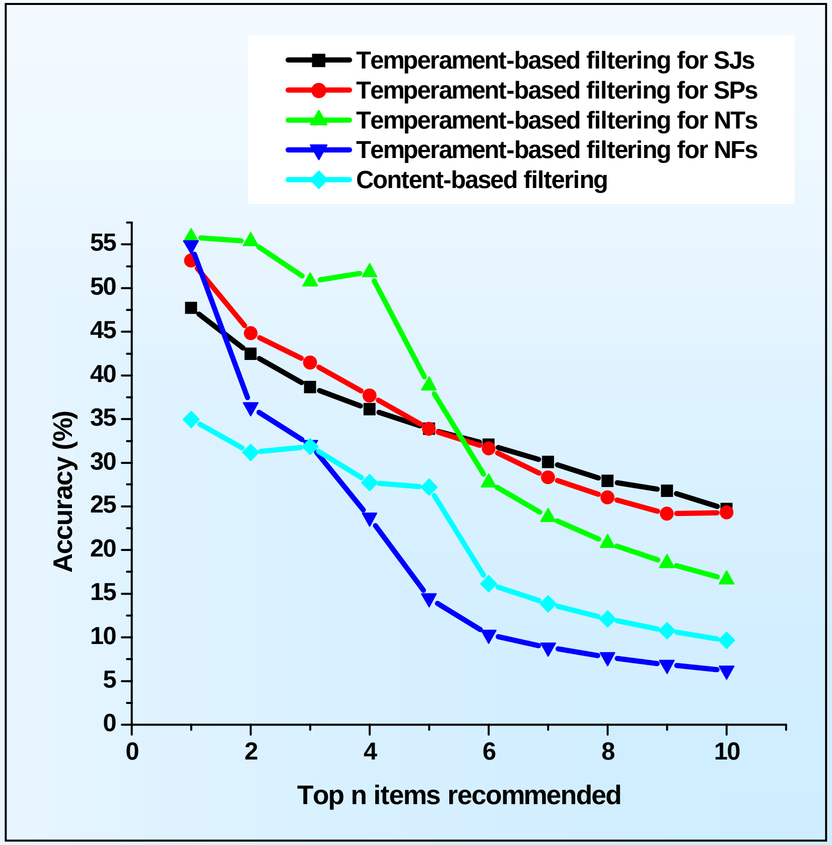
- **Accuracy of Recommendation (2000 documents, 20 tests)**

  (c) Given the user interest key terms

  (d) Given the user temperament and the user interest key terms



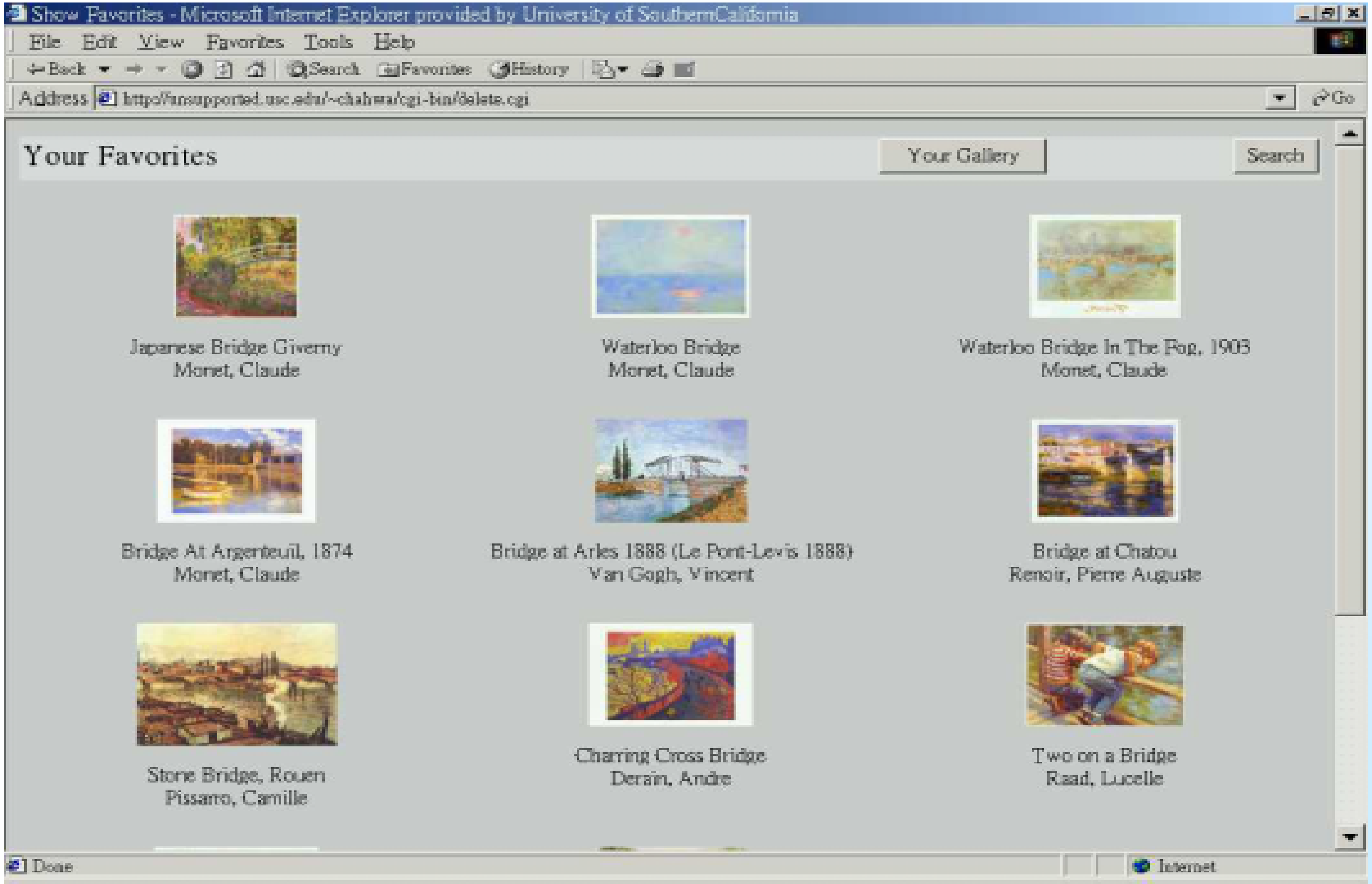
(c)



(d)

Experiment #2
An Art Image Collection of 2,000 Pictures

- **http://www.usc.edu/dept/cs/tfiltering/artworld**

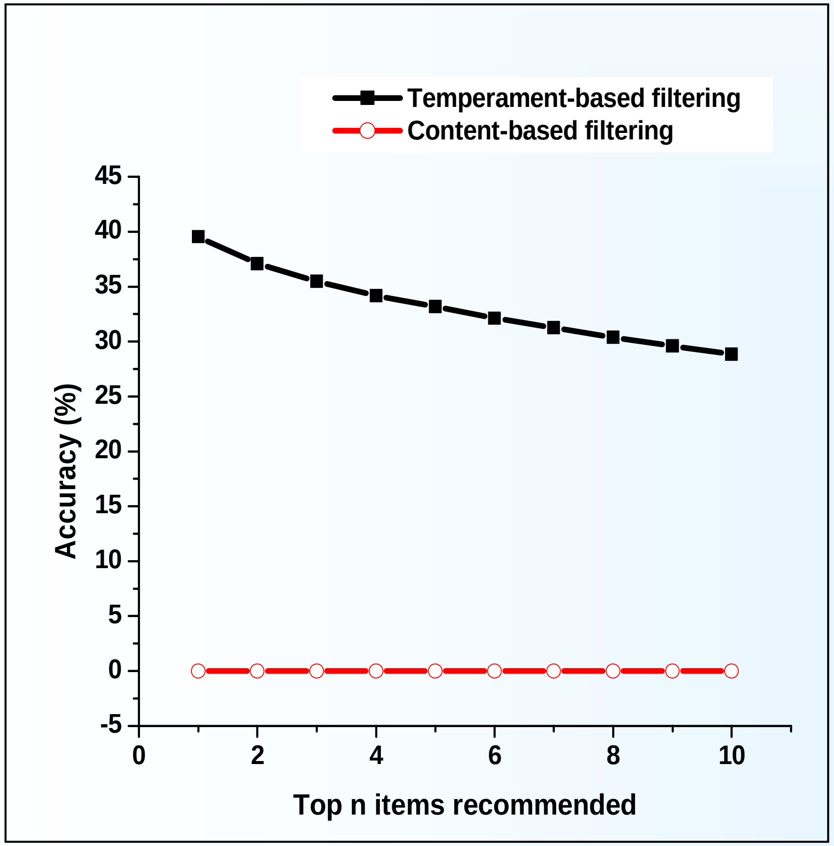


Experiment #2 - Cases (a) & (b)

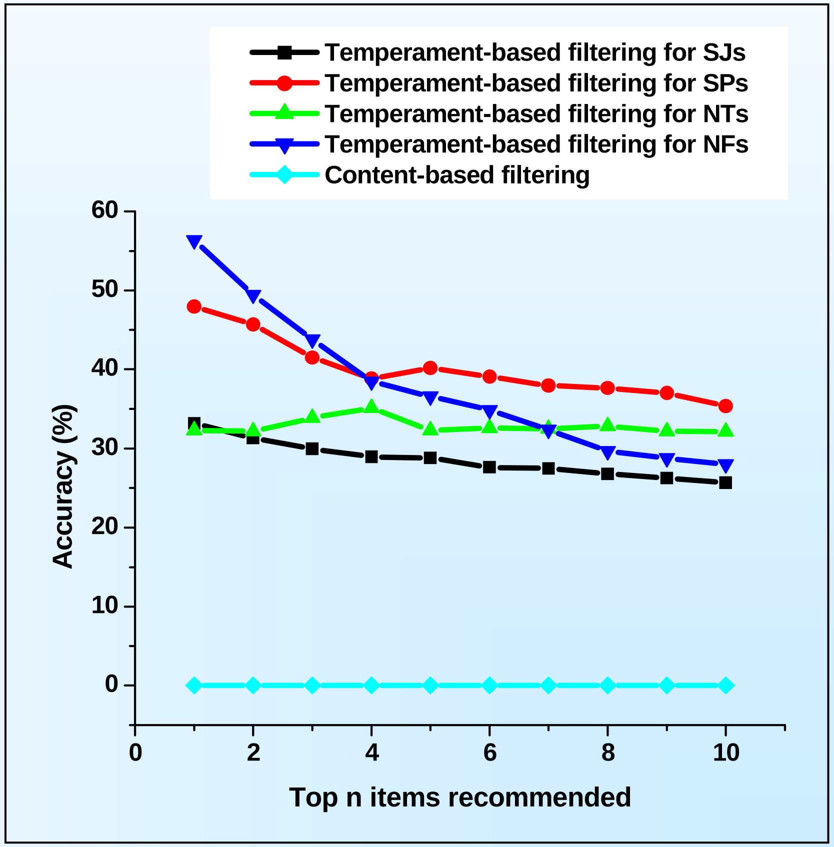
- **Accuracy of Recommendation (2000 art images, 20 tests)**

  (a) Users with unknown temperament and interest

  (b) Given the user temperament



(a)



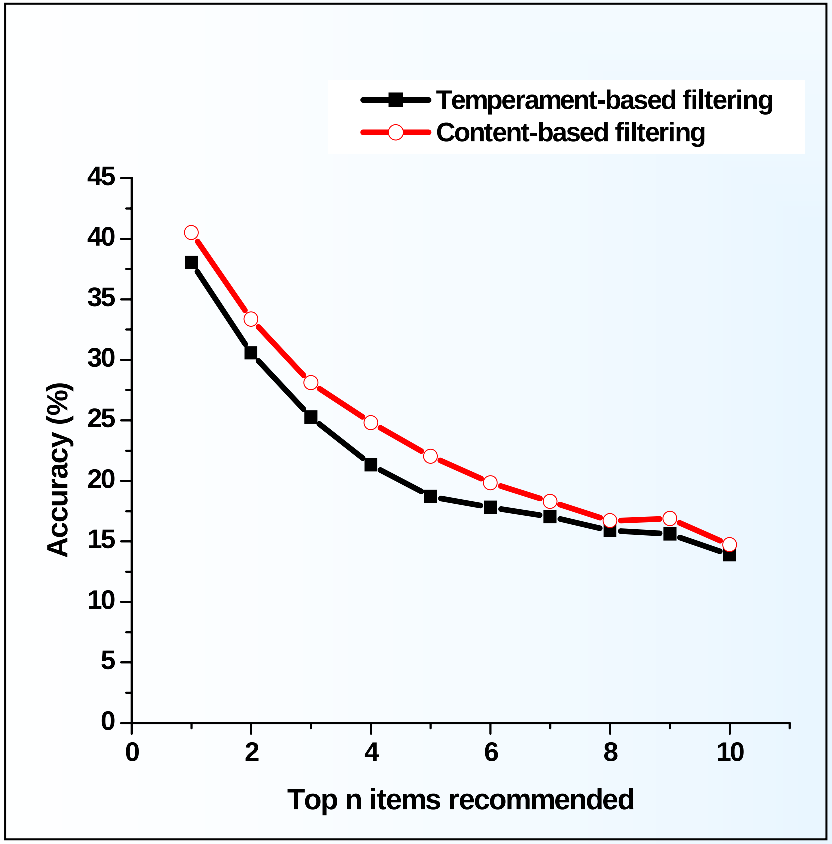
(b)

Experiment #2 - Cases (c) & (d)

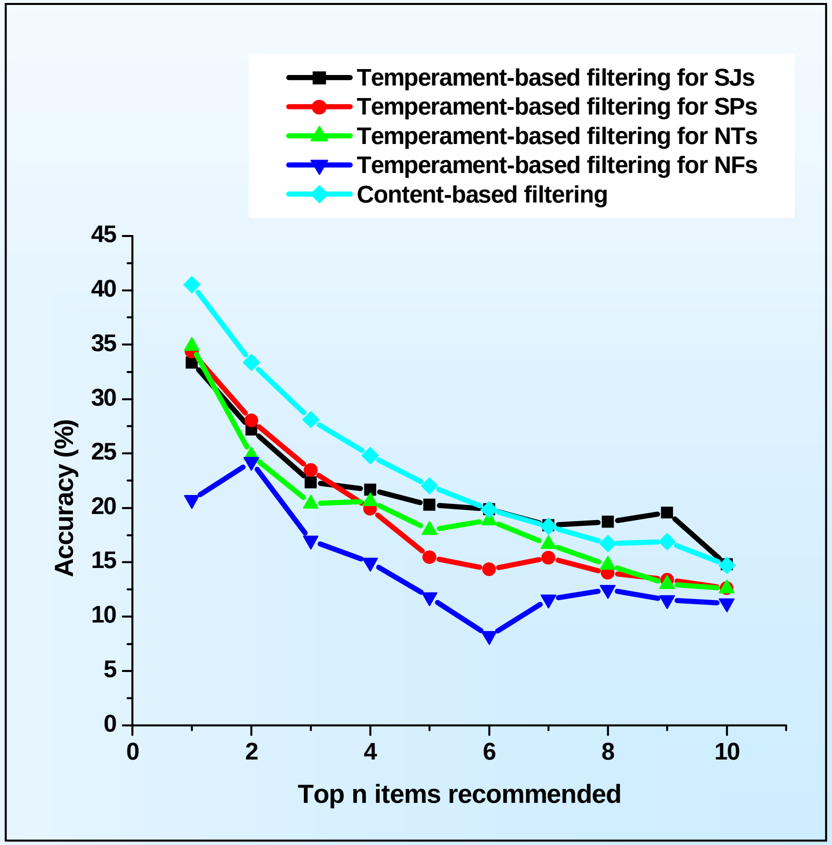
- **Accuracy of Recommendation (2000 art images, 20 tests)**

(c) Given the user interest key terms

(d) Given the user temperament and the user interest key terms



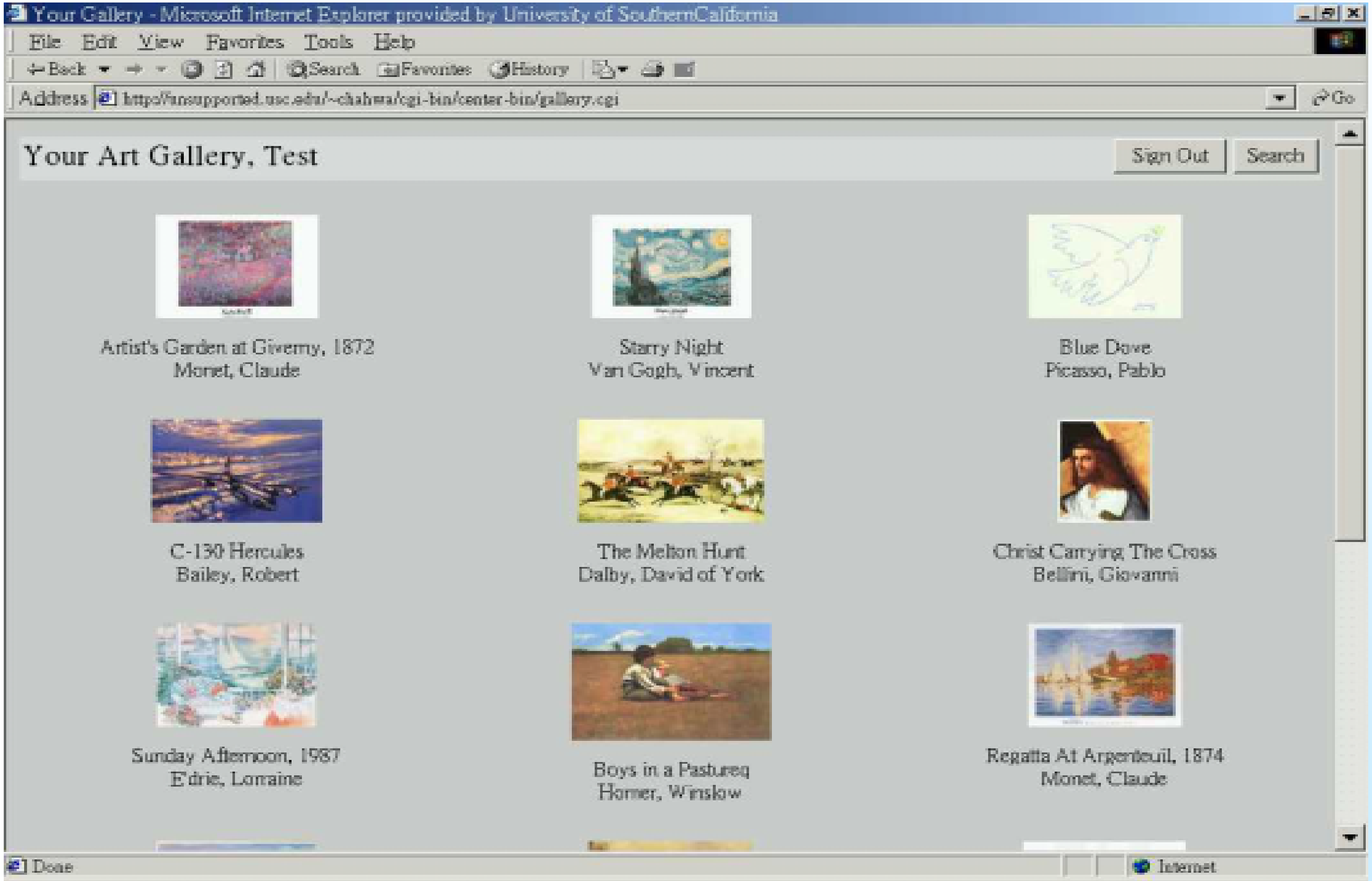
(c)



(d)

Experiment #3
An Art Image Collection of 100 Pictures

- **http://www.usc.edu/dept/cs/tfiltering/artcenter**

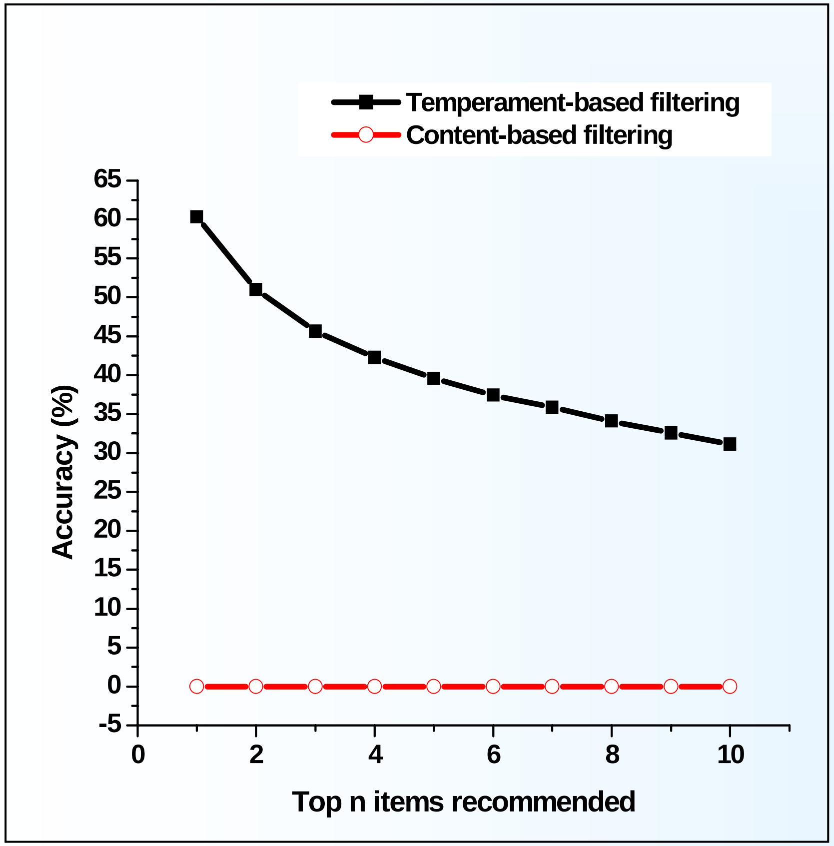


Experiment #3 - Cases (a) & (b)

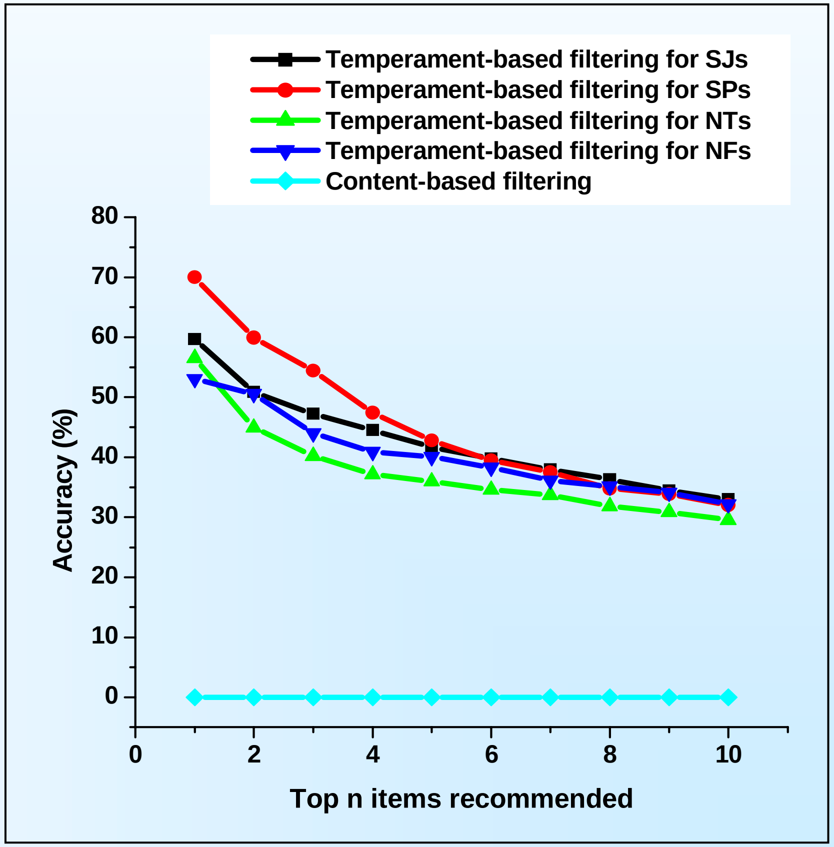
- **Accuracy of Recommendation (100 art images, 4 tests)**

  (a) Users with unknown temperament and interest

  (b) Given the user temperament



(a)



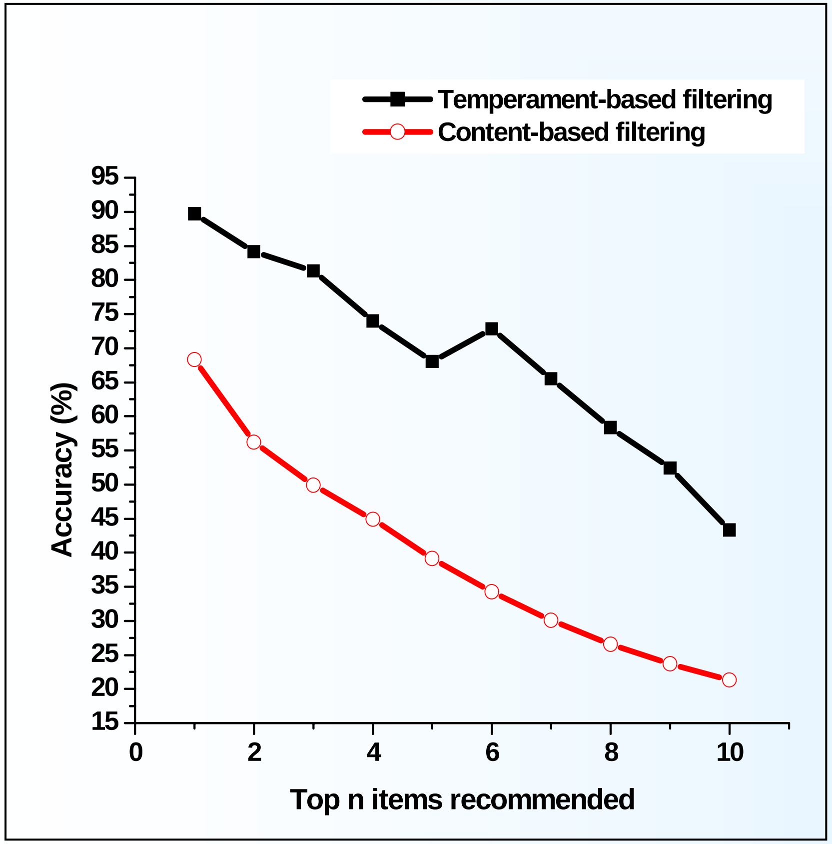
(b)

Experiment #3 - Cases (c) & (d)

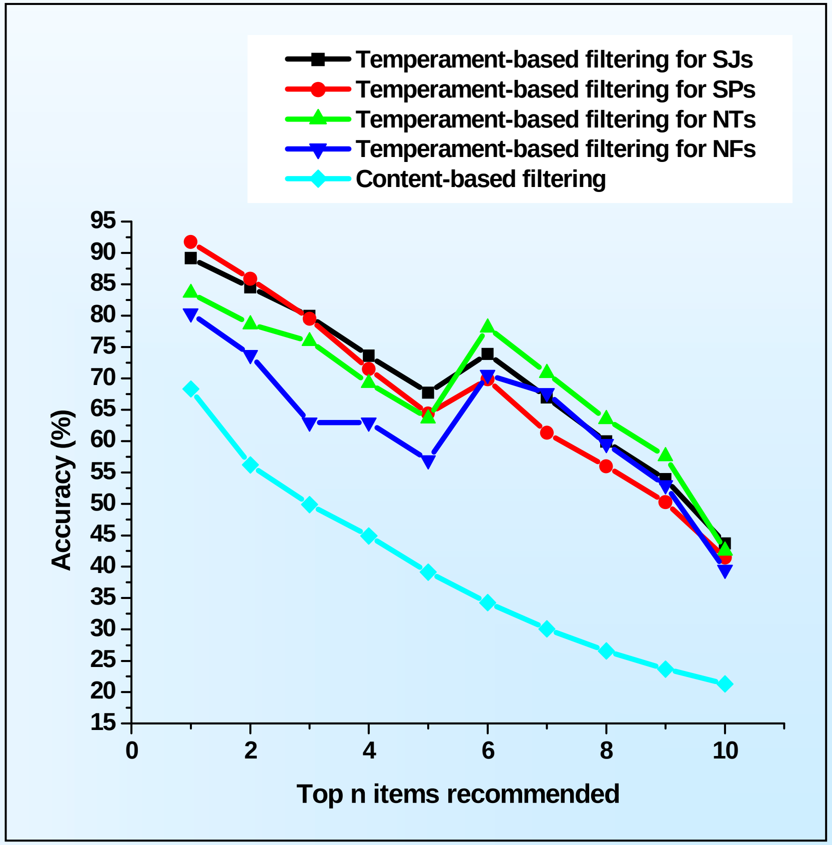
- **Accuracy of Recommendation (100 art images, 4 tests)**

  (c) Given the user interest key terms

  (d) Given the user temperament and the user interest key terms



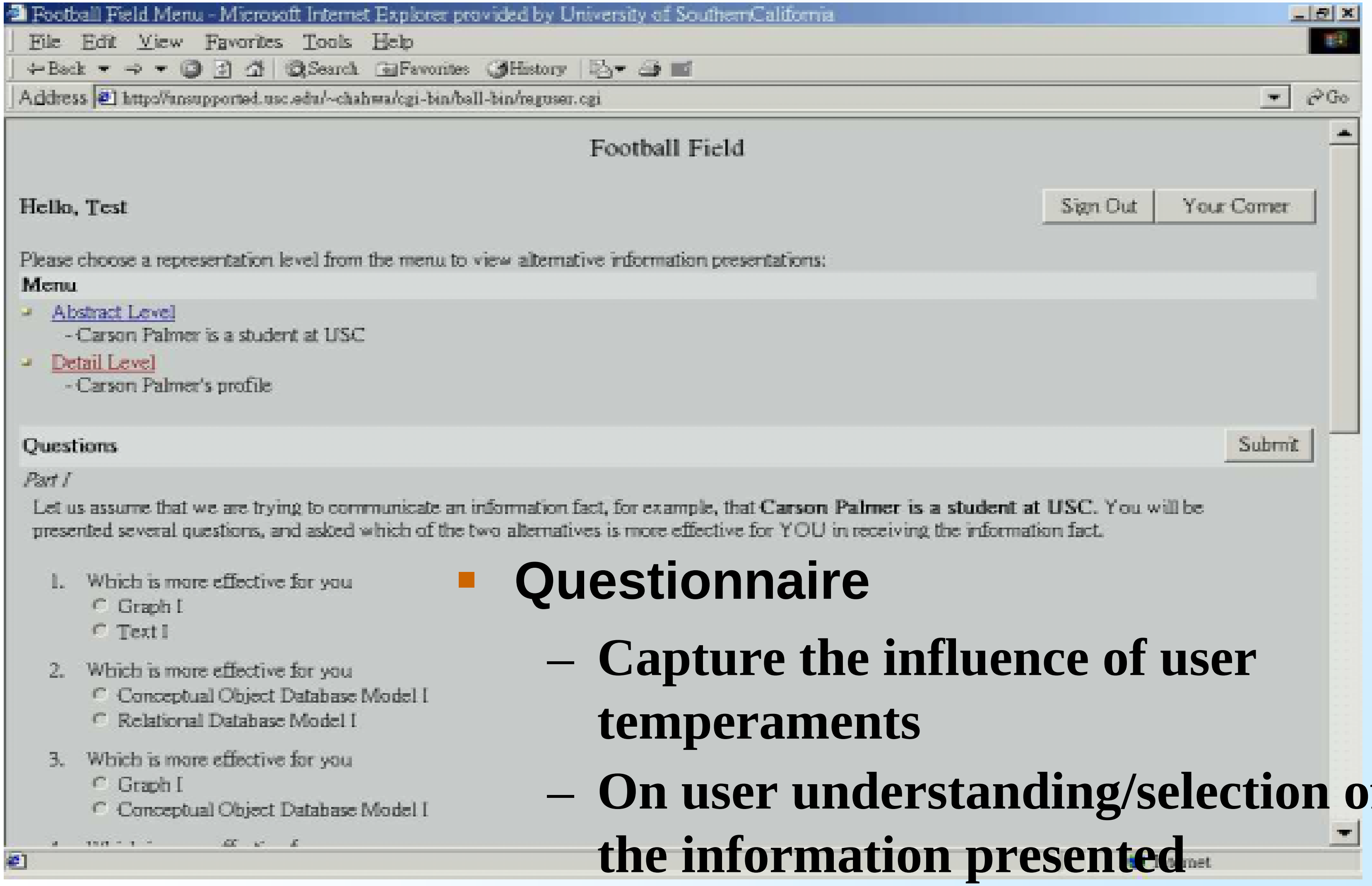
(c)



(d)

Experiment #4
Representation Styles

- **http://www.usc.edu/dept/cs/tfiltering/football**



- **Questionnaire**
  - Capture the influence of user temperaments
  - On user understanding/selection of the information presented

Experiment #4
Assumption & Representation

- **Assumption**
  - **Information units are randomly labeled with duplicated serial numbers/letters**
  - **No additional annotations are provided**
  - **Search by the key terms is not meaningful**

  ⇒Content-based filtering is not applicable
- **Representation**
  - **The information space could be static or dynamic**
  - **Two semantic levels of representation**

    Abstract: Carson Palmer is a student at USC

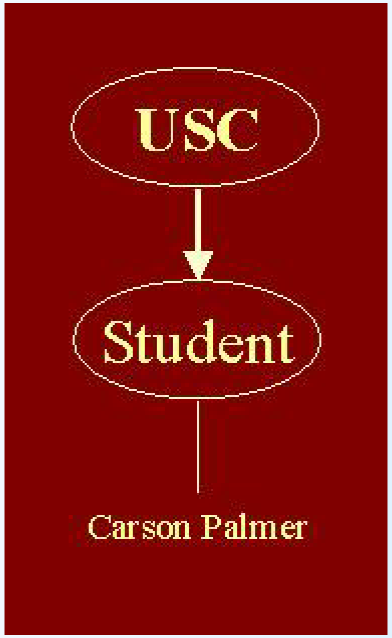
    Detailed: Carson Palmer's profile as a football player

Representation Styles - Abstract Level

- Carson Palmer is a student at the University of Southern California.

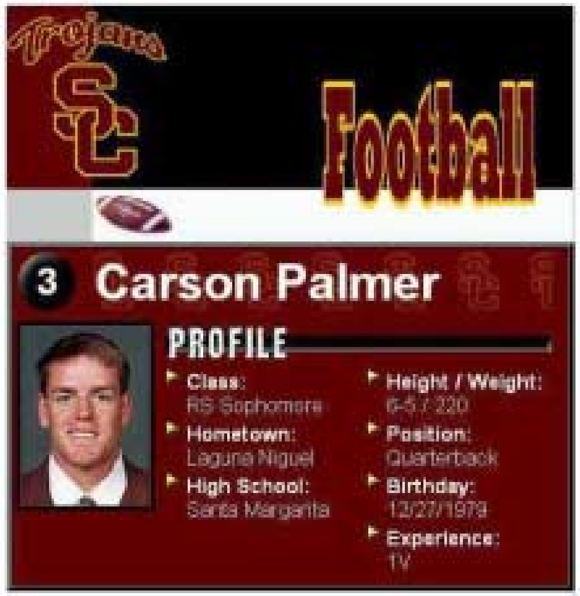
USC Students

Sultan Abdul-Malik
Shamsud-Din Abdul-Shaheed
Marcell Allmond
Kevin Arbet
Doyal Butler
Sunny Byrd
Chris Cash
Matt Cassel
.
.
.
Aaron Graham
Alex Holmes
Chahwa Lin
Zeke Moreno
Brennan Ochs
⇨ Carson Palmer
Kris Richard
Markus Steele
Zach Wilson



- **Clockwise**
  - Graph
  - Text
  - Concept object data model
  - Relational database model

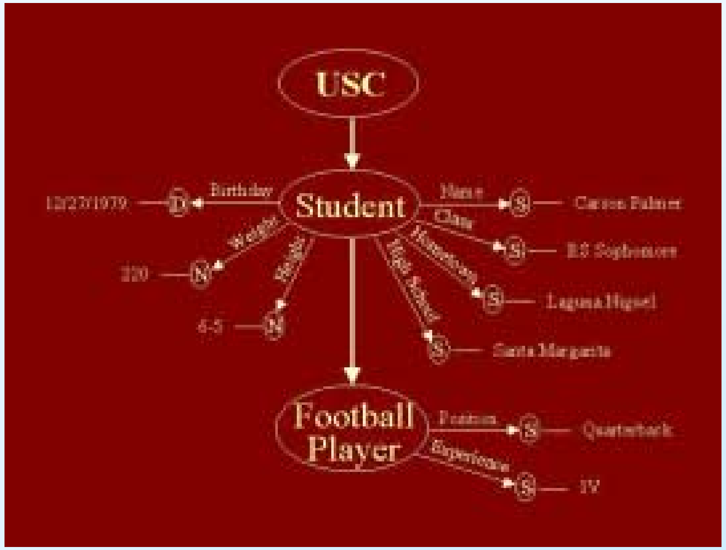
Representation Styles - Detail Level



- USC sophomore quarterback Carson Palmer was born on 12/27/79. The 6-foot-5, 220-pound Palmer with 1V experience comes from Laguna Niguna after graduated from Santa Margarita High School.

USC Football Roster

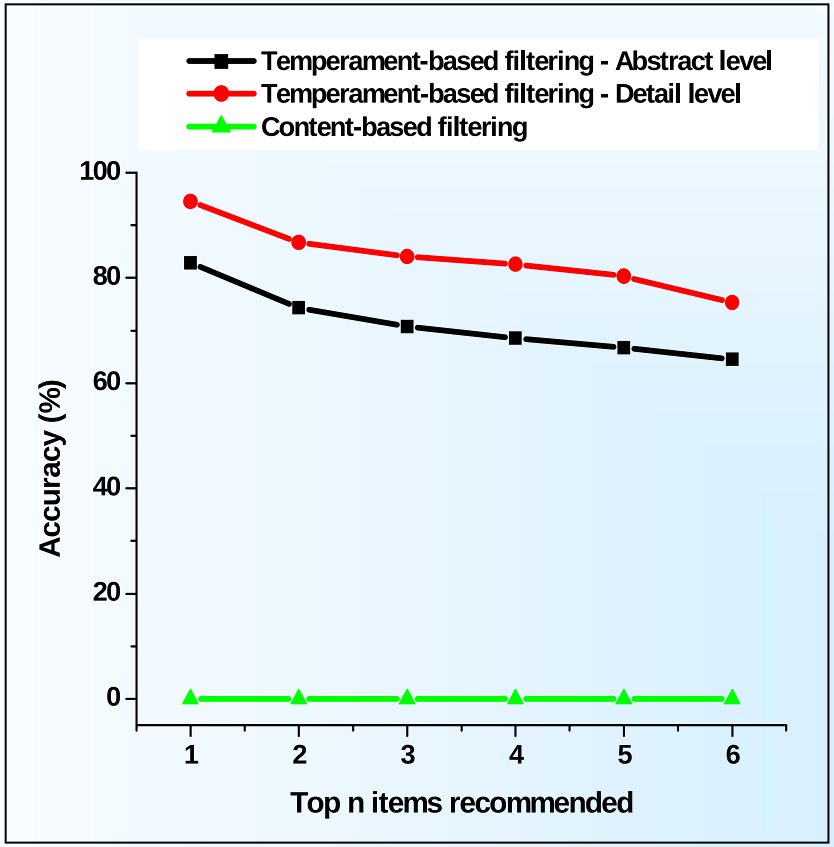
Name	Pos	Hgt	Wgt	Birthday	Cl	Exp	Hometown	High School
Sultan Abdul	DE	6-3	240	9/26/77	Sr	3V	Arcadia	Arcadia
Sham sud Abdul	DE-DT	6-4	250	10/12/77	Sr	3V	Los Angeles	Verbum Dei
Marcell Allmond	WR	6-1	190	3/28/81	So	1V	Anaheim	St. Paul
Kevin Arbet	S-CB	5-11	175	3/26/81	So	1V	Stockton	St. Mary's
Doyal Butler	TE	6-3	245	2/4/80	Jr	Jr	Tucson, AZ	Sabino
Aaron Graham	LB	6-1	225	6/12/81	So	1V	Bakersfield	Bakersfield
Chris Howard	TB	5-11	180	2/2/82	Fr	--	Los Angeles	Banning
Zeke Moreno	LB	6-3	245	10/10/78	Sr	3V	Chula Vista	Castle Park
Ifeanyi Ohalete	S	6-2	225	5/22/79	Sr	3V	Los Alamitos	Los Alamitos
Carson Palmer	**QB**	**6-5**	**220**	**12/27/79**	**So**	**1V**	**Laguna Niguel**	**Santa Margarita**
Kris Richard	CB	6-0	180	10/28/78	Jr	2V	Carson	Serra
Markus Steele	LB	6-3	220	7/24/79	Sr	1V	Long Beach	Chanel
Zach Wilson	OG-OT	6-5	315	10/14/79	So	1V	Bellflower	Mayfair



Experiment #4.1
Static Information Space

- **Accuracy of Recommendation (2000 simulated users, 20 tests)**

(a) Users with unknown temperament and interest



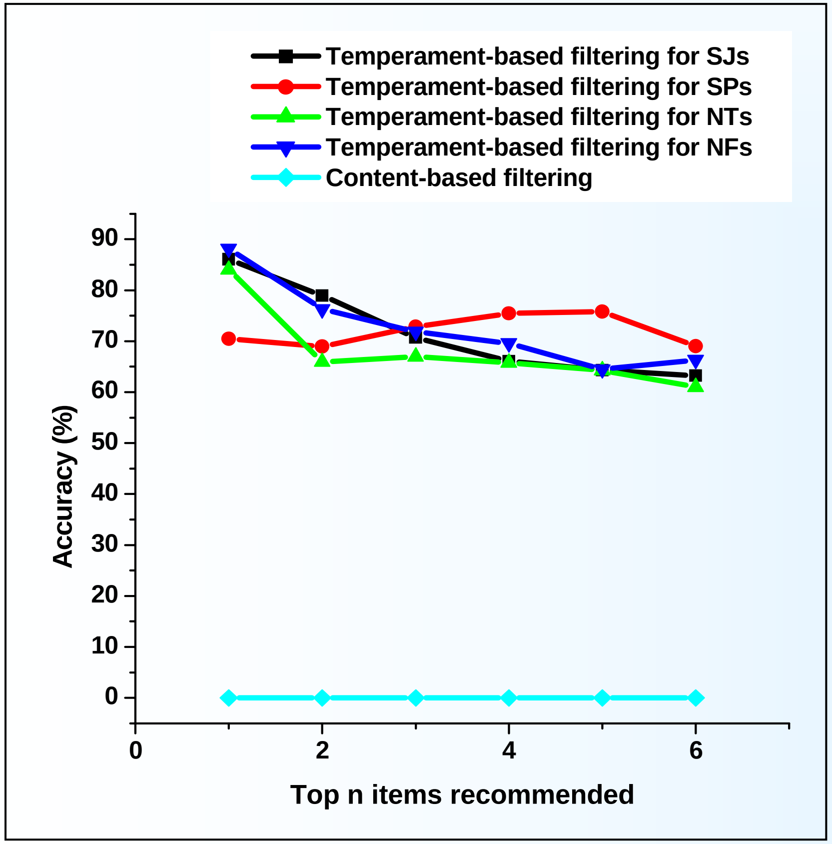
(a)

Experiment #4.1
Static Information Space (cont.)

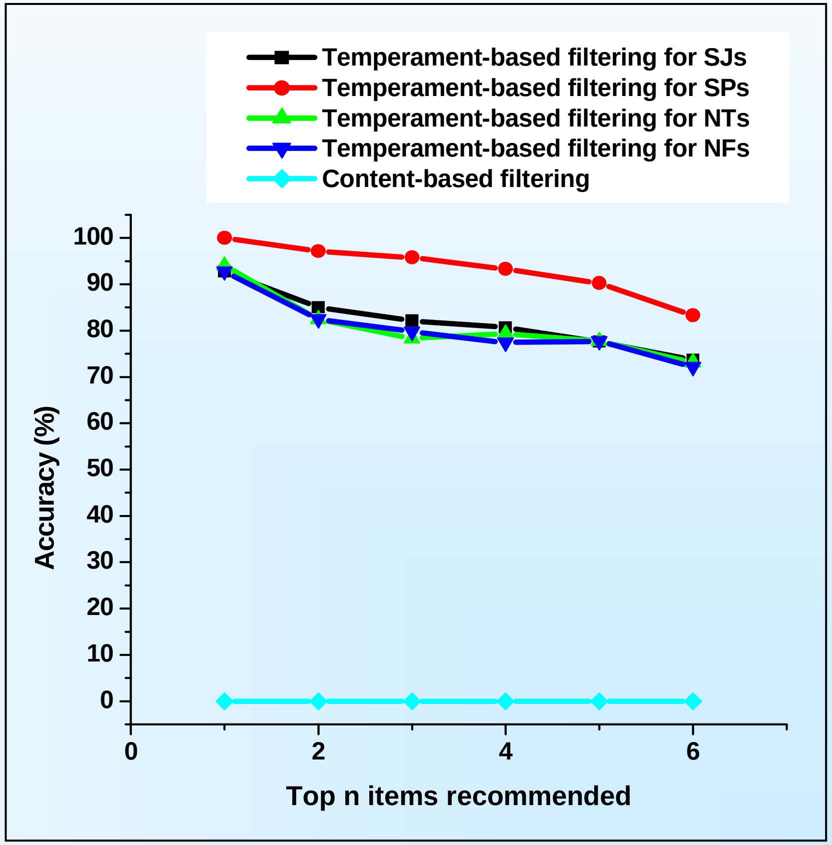
- **Accuracy of Recommendation (2000 simulated users, 20 tests)**

(b) Given the user temperament - Abstract level

(c) Given the user temperament - Detail level



(b)

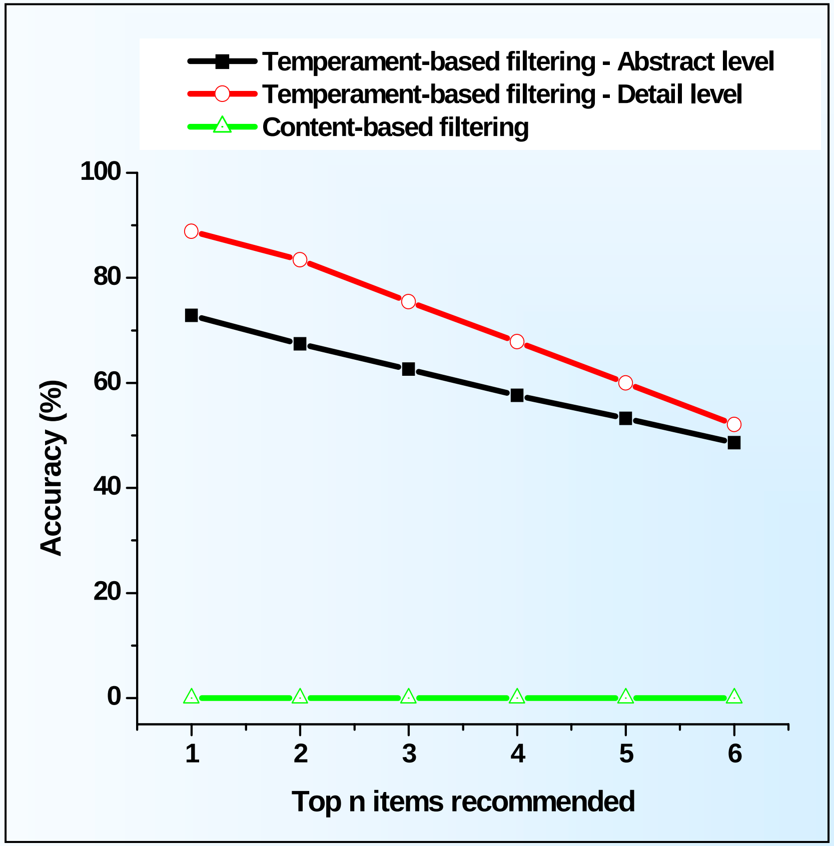


(c)

Experiment #4.2
Dynamic Information Space

- **Accuracy of Recommendation (12 items/6 questions, 20 tests)**

(a) Users with unknown temperament and interest



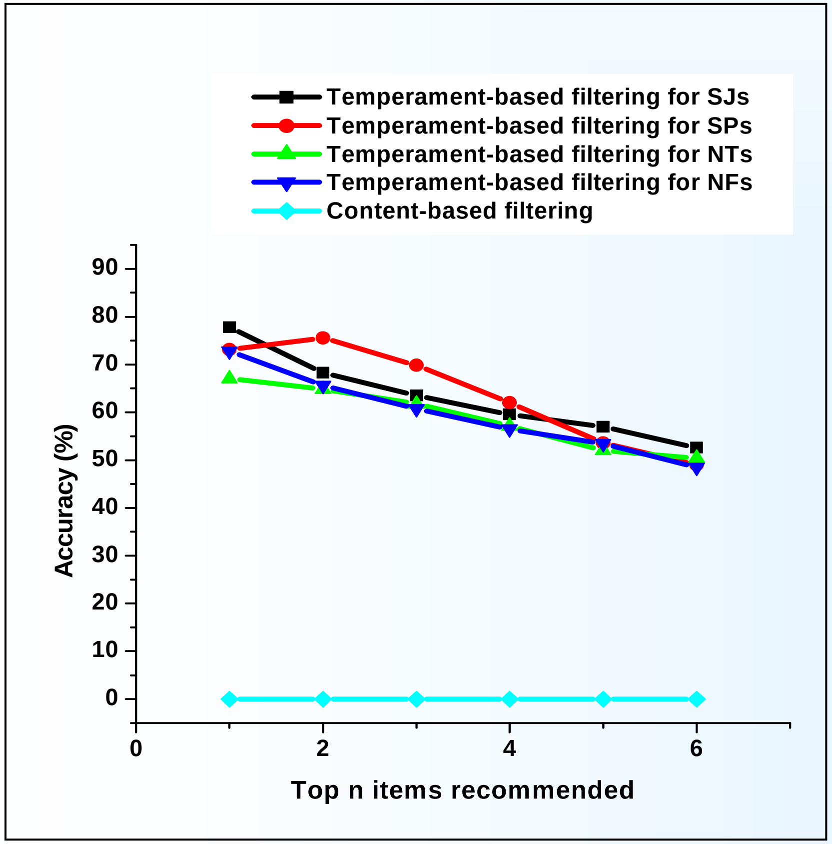
(a)

Experiment #4.2
Dynamic Information Space (cont.)

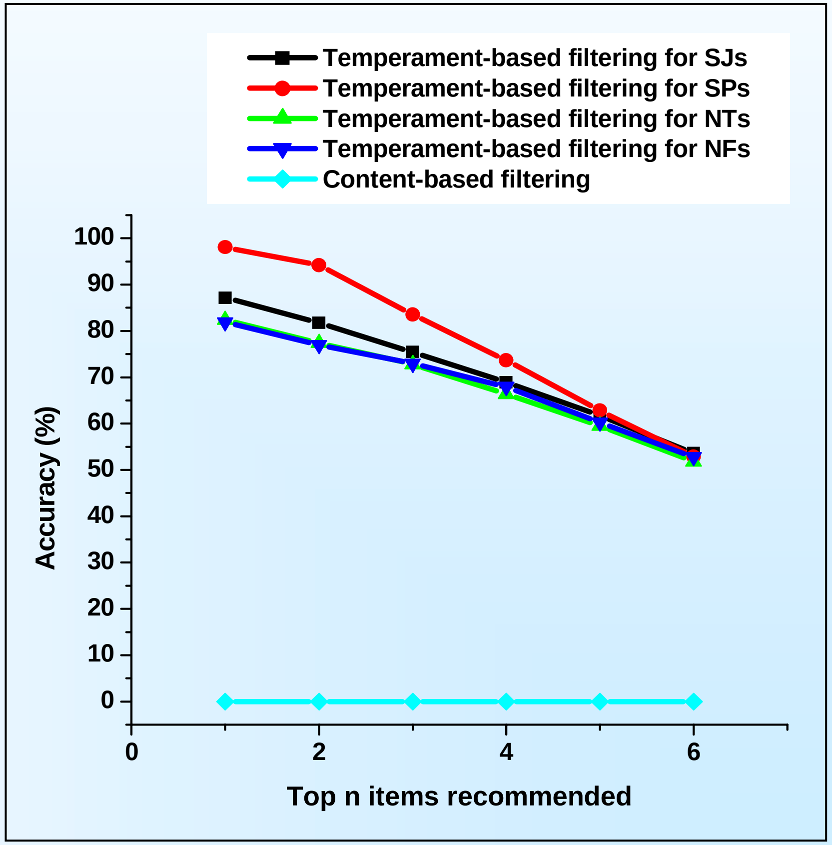
- **Accuracy of Recommendation (12 items/6 questions, 20 tests)**

  (b) Given the user temperament - Abstract level

  (c) Given the user temperament - Detail level



(b)



(c)

Popularity Similarity Measures

- **Level of importance of both popularity and similarity**
  - **Possible equations**

$$PopSim(V_c, V_k) = e_c + Sim(V_c, V_k)$$

and

$$PopSim(V_c, V_k) = e_c Sim(V_c, V_k)$$

  - **Classification function**

$$(S_{target}, C_{target}) = \arg\max_{s \in S, c \in C_s} PopSim(V_c, V_k)$$

- **Examination**
  - **Two possible equations and cosine similarity measure**
  - **Three collections of experiments #1, #2, and #3**
  - **Four user conditions**

Comparison of Popularity Similarity Measures 2,000 Documents - Cases (a) & (b)

- **Accuracy of Recommendation (20 tests)**

  (a) Users with unknown temperament and interest.

  (b) Given the user temperament.

Top n items	Accuracy of Recommendation, %		
	$e_c + Sim(V_c, V_k)$	$e_c\, Sim(V_c, V_k)$	$Sim(V_c, V_k)$
1	41	39	37
2	36	35	31
3	35	32	29
4	34	30	24
5	33	28	25
6	32	26	22
7	31	24	24
8	31	22	26
9	30	20	23
10	30	19	19

(a)

Top n items	Accuracy of Recommendation, %											
	$e_c + Sim(V_c, V_k)$				$e_c\, Sim(V_c, V_k)$				$Sim(V_c, V_k)$			
	SJ	SP	NT	NF	SJ	SP	NT	NF	SJ	SP	NT	NF
1	34	52	42	44	34	52	42	39	33	52	45	36
2	29	44	41	44	30	44	40	40	29	45	42	40
3	27	42	40	41	28	42	39	38	27	42	42	35
4	27	43	41	36	26	43	40	36	26	42	41	34
5	26	42	40	35	26	41	40	33	25	40	41	32
6	26	40	39	33	25	40	39	31	25	39	40	29
7	25	40	38	31	24	39	40	28	24	37	39	28
8	25	39	39	28	24	39	39	28	24	36	39	29
9	25	39	39	26	24	38	38	28	23	35	37	29
10	25	39	39	24	23	37	38	27	22	34	36	29

(b)

Comparison of Popularity Similarity Measures 2,000 Documents - Cases (c) & (d)

- **Accuracy of Recommendation (20 tests)**

(c) Given the user interest key terms.

(d) Given the user temperament and the user interest key terms.

Top n items	Accuracy of Recommendation, % $e_c + Sim(V_c, V_k)$	$e_c\, Sim(V_c, V_k)$	$Sim(V_c, V_k)$
1	54	35	47
2	45	31	44
3	41	31	46
4	37	30	39
5	33	27	32
6	32	23	26
7	30	23	22
8	27	19	20
9	26	18	17
10	25	16	16

(c)

Top n items	Accuracy of Recommendation, %											
	$e_c + Sim(V_c, V_k)$				$e_c\, Sim(V_c, V_k)$				$Sim(V_c, V_k)$			
	SJ	SP	NT	NF	SJ	SP	NT	NF	SJ	SP	NT	NF
1	48	53	56	55	41	43	54	45	43	45	46	46
2	42	45	55	36	34	43	52	33	40	47	43	29
3	39	41	51	32	30	39	49	26	46	49	52	24
4	36	38	52	24	29	37	47	23	36	37	39	19
5	34	34	39	15	29	34	44	20	29	29	31	15
6	32	32	28	10	25	30	42	18	24	25	26	13
7	30	28	24	9	24	30	40	17	21	21	22	11
8	28	26	21	8	19	24	32	16	18	18	20	10
9	27	24	19	7	17	19	26	14	16	16	17	8
10	25	24	17	6	16	17	24	13	14	15	16	8

(d)

Comparison of Popularity Similarity Measures 2,000 Art Images - Cases (a) & (b)

- **Accuracy of Recommendation (20 tests)**

  (a) Users with unknown temperament and interest.

  (b) Given the user temperament.

Top n items	Accuracy of Recommendation, %		
	$e_c + Sim(V_c, V_k)$	$e_c\,Sim(V_c, V_k)$	$Sim(V_c, V_k)$
1	40	29	21
2	37	25	18
3	35	23	12
4	34	20	12
5	33	17	9
6	32	16	8
7	31	14	7
8	30	13	6
9	30	13	5
10	29	13	5

(a)

Top n items	Accuracy of Recommendation, %											
	$e_c + Sim(V_c, V_k)$				$e_c\,Sim(V_c, V_k)$				$Sim(V_c, V_k)$			
	SJ	SP	NT	NF	SJ	SP	NT	NF	SJ	SP	NT	NF
1	33	48	32	56	31	52	24	43	29	49	21	31
2	31	46	32	49	31	45	27	34	25	42	21	26
3	30	41	34	44	29	39	28	29	23	37	25	21
4	29	39	35	38	27	38	25	27	21	35	24	20
5	29	40	32	37	25	37	24	24	20	35	25	19
6	28	39	33	35	23	38	22	22	18	34	28	16
7	27	38	33	32	22	37	20	20	19	33	27	15
8	27	38	33	30	21	36	19	19	18	33	26	15
9	26	37	32	29	19	35	18	18	21	32	27	15
10	26	35	32	28	18	33	15	18	16	29	16	13

(b)

Comparison of Popularity Similarity Measures 2,000 Art Images - Cases (c) & (d)

- **Accuracy of Recommendation (20 tests)**

(c) Given the user interest key terms.

(d) Given the user temperament and the user interest key terms.

Top n items	Accuracy of Recommendation, %		
	$e_c + Sim(V_c, V_k)$	$e_c\, Sim(V_c, V_k)$	$Sim(V_c, V_k)$
1	38	37	36
2	31	34	30
3	25	30	25
4	21	27	22
5	19	24	19
6	18	23	17
7	17	21	16
8	16	20	15
9	16	18	15
10	14	17	14

(c)

Top n items	Accuracy of Recommendation, %											
	$e_c + Sim(V_c, V_k)$				$e_c\, Sim(V_c, V_k)$				$Sim(V_c, V_k)$			
	SJ	SP	NT	NF	SJ	SP	NT	NF	SJ	SP	NT	NF
1	33	34	35	21	30	46	27	30	32	35	27	25
2	27	28	25	24	26	39	20	27	28	26	22	22
3	22	23	20	17	25	34	20	26	23	22	21	20
4	22	20	21	15	24	30	23	33	20	19	18	19
5	20	15	18	12	21	27	24	35	19	16	17	16
6	20	14	19	8	21	26	33	35	17	9	16	14
7	18	15	17	12	21	23	35	38	16	8	16	14
8	19	14	15	12	20	21	31	36	16	6	21	11
9	20	13	13	12	19	20	28	40	17	1	19	9
10	15	13	13	11	15	17	14	18	15	1	13	8

(d)

Comparison of Popularity Similarity Measures 100 Art Images - Cases (a) & (b)

- **Accuracy of Recommendation**
  (4 tests)

  (a) Users with unknown temperament and interest.

  (b) Given the user temperament.

Top n items	Accuracy of Recommendation, % $e_c + Sim(V_c, V_k)$	$e_c\,Sim(V_c, V_k)$	$Sim(V_c, V_k)$
1	60	0	0
2	51	0	0
3	46	0	0
4	42	0	0
5	40	0	0
6	37	0	0
7	36	0	0
8	34	0	0
9	33	0	0
10	31	0	0

(a)

Top n items	Accuracy of Recommendation, %											
	$e_c + Sim(V_c, V_k)$				$e_c\,Sim(V_c, V_k)$				$Sim(V_c, V_k)$			
	SJ	SP	NT	NF	SJ	SP	NT	NF	SJ	SP	NT	NF
1	60	70	57	53	49	56	41	53	46	55	48	47
2	51	60	45	51	42	61	32	51	38	44	39	41
3	47	54	40	44	50	53	28	42	49	35	47	39
4	44	47	37	41	44	45	39	41	45	33	44	36
5	42	43	36	40	41	42	37	40	42	29	40	33
6	40	39	35	38	37	38	36	37	41	27	38	31
7	38	38	34	36	35	35	35	36	40	28	38	29
8	36	35	32	35	33	34	34	33	38	30	35	28
9	34	34	31	34	38	32	33	32	37	30	34	25
10	33	32	30	32	29	28	26	31	36	24	28	29

(b)

Comparison of Popularity Similarity Measurs 100 Art Images - Cases (c) & (d)

- **Accuracy of Recommendation (4 tests)**

  (c) Given the user interest key terms.

  (d) Given the user temperament and the user interest key terms.

Top n items	Accuracy of Recommendation, %		
	$e_c + Sim(V_c, V_k)$	$e_c\, Sim(V_c, V_k)$	$Sim(V_c, V_k)$
1	90	65	54
2	84	57	45
3	81	47	38
4	74	45	31
5	68	38	25
6	73	37	21
7	66	32	18
8	58	28	15
9	52	25	13
10	43	22	14

(c)

Top n items	Accuracy of Recommendation, %											
	$e_c + Sim(V_c, V_k)$				$e_c\, Sim(V_c, V_k)$				$Sim(V_c, V_k)$			
	SJ	SP	NT	NF	SJ	SP	NT	NF	SJ	SP	NT	NF
1	89	92	84	80	64	51	53	53	43	48	54	45
2	85	86	79	74	65	56	53	43	48	34	52	41
3	80	80	76	63	60	50	38	41	40	29	43	41
4	74	71	69	63	66	53	40	33	46	30	47	35
5	68	64	64	57	55	48	33	30	37	33	39	29
6	74	70	78	71	50	45	33	29	32	28	33	24
7	67	61	71	68	44	38	32	25	27	24	29	22
8	60	56	64	60	38	36	28	22	24	17	25	19
9	54	50	58	53	34	32	25	19	21	15	22	17
10	44	41	43	39	29	26	23	17	13	16	20	17

(d)

Hypothesis & Goal

- **The accuracy of an information recommendation system can be significantly improved by employing user temperament for filtering and customization**
- **To characterize the information space by taking human factors into consideration and devise a new filtering mechanism to provide a better information recommendation service**

Anticipated Contributions

- **New general approach to customized information selection**
- **Incorporate human factors (temperament) into the adaptive information recommendation process**
- **Analysis of the inherent interrelated patterns between user temperaments and user interests**
- **The accuracy of an information recommendation system can be significantly improved**
- **Basis for exploring other human factors**

Direct Research Extension

- **Focus on temperament**
  - **Other characteristics, human factors**
  - **E.g., gender, age, education level, experience with system, user demographics**
- **Multi-level scale of rating**
  - **Not just like and dislike, but degrees**
- **Extend user-studies beyond coherent "adults only" population**