Mining User Navigation Patterns for Personalizing Topic Directories

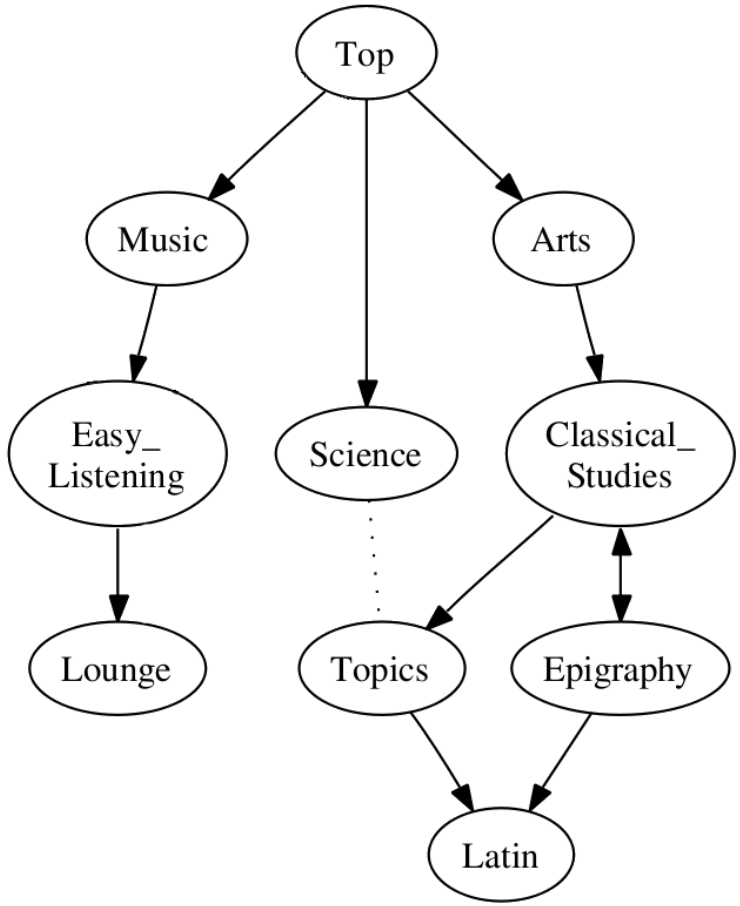
Theodore Dalamagas, Panagiotis Bouros, Theodore Galanis, Magdalini Eirinaki and Timos Sellis

Panagiotis Bouros
Knowledge and Database Systems Lab
School of Electrical and Computer Engineering
National Technical University of Athens, Greece

Outline

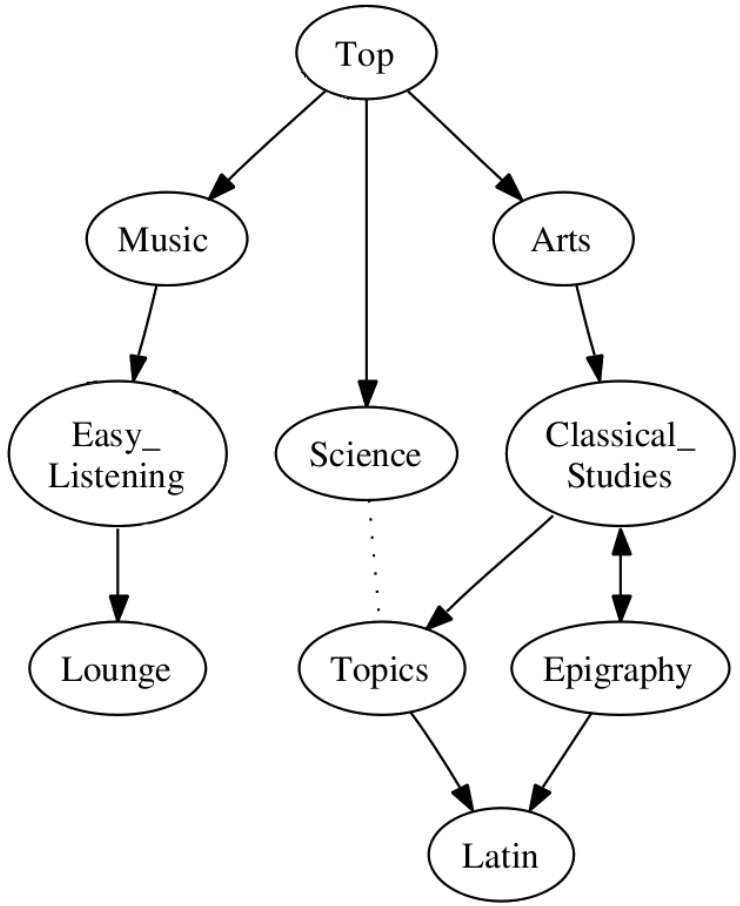
Introduction

- Topic directories, popular means of organizing web resources
  - Hierarchical organization of thematic categories
- As search "tools"
  - Narrowing search from broad topics to specific ones, e.g. `Arts` to `Classical_Studies`
  - Support keyword search



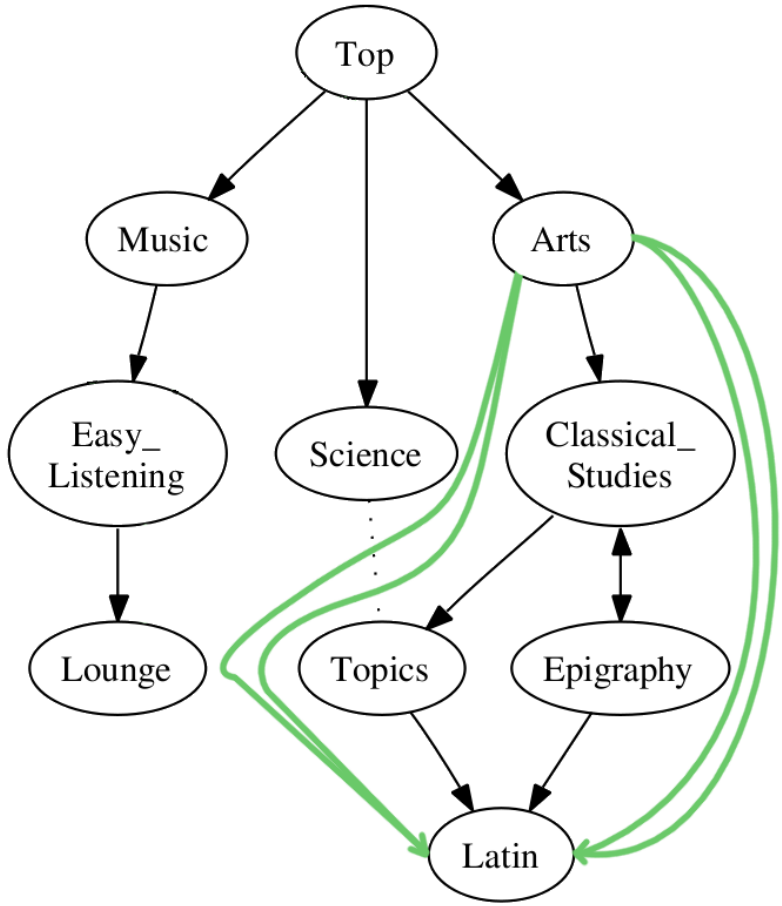
Introduction

- Topic directories, popular means of organizing web resources
  - Hierarchical organization of thematic categories
- As search "tools"
  - Narrowing search from broad topics to specific ones, e.g. `Arts` to `Classical_Studies`
  - Support keyword search
- Need for personalization
  - Huge amount of web resources
  - Growing diversity of web data sources
  - Heterogeneity of user communities



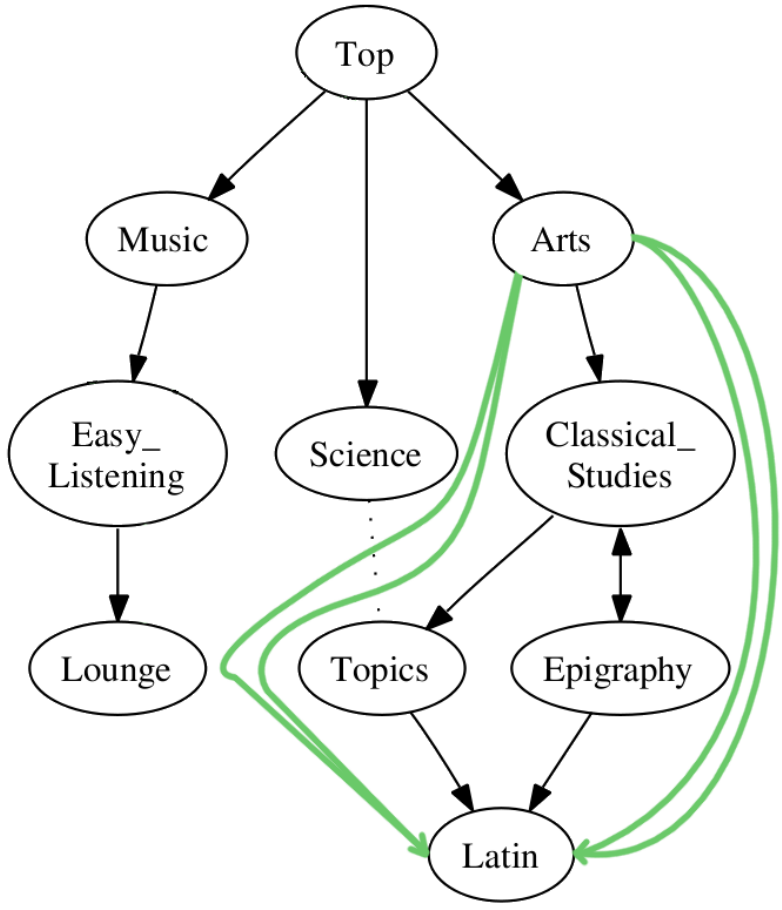
Introduction

- Topic directories, popular means of organizing web resources
  - Hierarchical organization of thematic categories
- As search "tools"
  - Narrowing search from broad topics to specific ones, e.g. `Arts` to `Classical_Studies`
  - Support keyword search
- Need for personalization
  - Huge amount of web resources
  - Growing diversity of web data sources
  - Heterogeneity of user communities
- Personalizing topic directories
  - Provide a "view" of topic directory tailored to user needs
  - Bypass topics not tailored to user needs



Introduction

- Topic directories, popular means of organizing web resources
  - Hierarchical organization of thematic categories
- As search "tools"
  - Narrowing search from broad topics to specific ones, e.g. `Arts` to `Classical_Studies`
  - Support keyword search
- Need for personalization
  - Huge amount of web resources
  - Growing diversity of web data sources
  - Heterogeneity of user communities
- Personalizing topic directories
  - Provide a "view" of topic directory tailored to user needs
  - Bypass topics not tailored to user needs



- Provide direct link from `Arts` to `Latin` for users interested in `Latin`

Contribution in brief

- Methods to personalize topic directories
  - Provide topic directory views
  - Views are based on users navigation history - behaviour

Contribution in brief

- Methods to personalize topic directories
  - Provide topic directory views
  - Views are based on users navigation history - behaviour
- Personalization
  - Involves adding new links called shortcuts in the directory
    - Offline (static shortcuts) - presented to groups of users with similar navigation behaviour
    - Online (dynamic shortcuts) - presented to each individual user
  - Shortcuts help users to easily reach topics tailored to their needs, while bypass others
    - `Arts`$\rightarrow$`Latin`
  - Personalization is based on a set of mining tasks
    - e.g., identifying interest groups, users with certain type of behaviour, etc. (see later slides)

Contribution in brief

- Methods to personalize topic directories
  - Provide topic directory views
  - Views are based on users navigation history - behaviour
- Personalization
  - Involves adding new links called shortcuts in the directory
    - Offline (static shortcuts) - presented to groups of users with similar navigation behaviour
    - Online (dynamic shortcuts) - presented to each individual user
  - Shortcuts help users to easily reach topics tailored to their needs, while bypass others
    - `Arts`$\rightarrow$`Latin`
  - Personalization is based on a set of mining tasks
    - e.g., identifying interest groups, users with certain type of behaviour, etc. (see later slides)
- Experimental evaluation of both mining and personalization tasks

Outline

Modelling topic directories

Topic directory

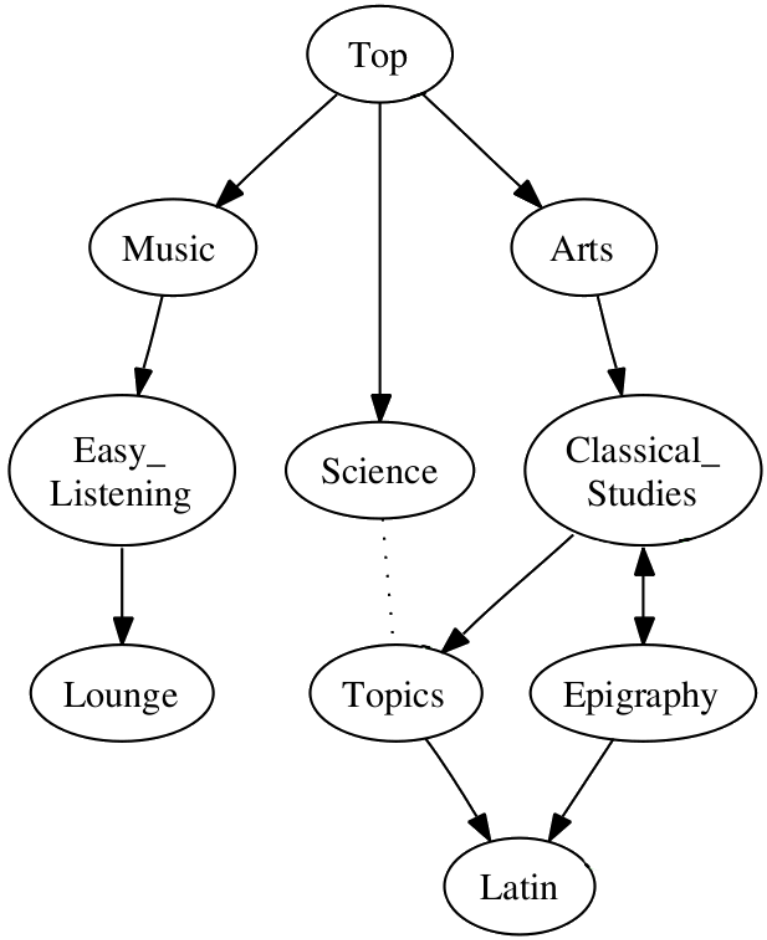
- Hierarchical organization of thematic categories
- Categories contain resources, i.e. links to other pages
- Subcategories narrow content of broad categories
- Related categories contain similar resources
- Directory graph

Modelling topic directories

Topic directory

- Hierarchical organization of thematic categories
- Categories contain resources, i.e. links to other pages
- Subcategories narrow content of broad categories
- Related categories contain similar resources
- Directory graph

Example



Modelling topic directories

Topic directory

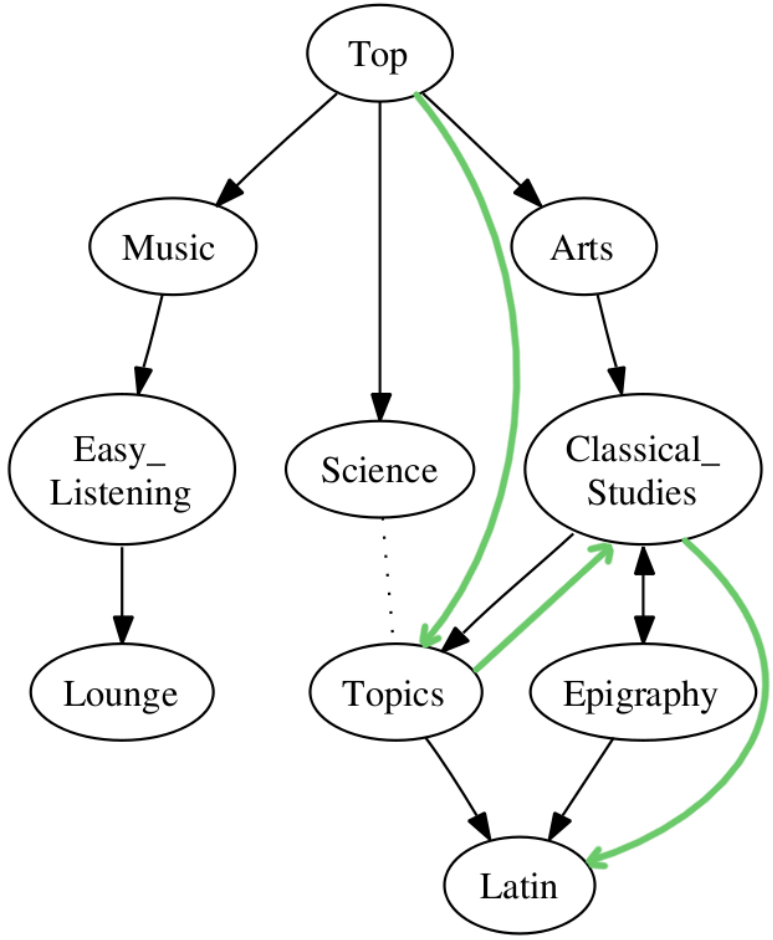
- Hierarchical organization of thematic categories
- Categories contain resources, i.e. links to other pages
- Subcategories narrow content of broad categories
- Related categories contain similar resources
- Directory graph

Navigation pattern

- Sequence of categories during session
- Navigation behaviour of users for reaching more than one topic
- Multiple occurrences of same categories, i.e. back and forth

Example

{Top,Arts,Classical_Studies,Topics,
Classical_Studies,Epigraphy,Latin}



Outline

Overview of mining tasks

- Identifying interest groups
  - Users with similar navigation behaviour - interests
  - Clustering user navigation patterns
  - Navigation patterns similarity

Overview of mining tasks

- Identifying interest groups
  - Users with similar navigation behaviour - interests
  - Clustering user navigation patterns
  - Navigation patterns similarity
- Identifying indecisive users
  - "Back and forth" to same categories

Overview of mining tasks

- Identifying interest groups
  - Users with similar navigation behaviour - interests
  - Clustering user navigation patterns
  - Navigation patterns similarity
- Identifying indecisive users
  - "Back and forth" to same categories
- Mining (L-)popular categories & sequential navigation (L-)subpatterns
  - Popular categories, i.e., frequently visited
  - (L-)popular categories, i.e., contain frequently selected resources
  - Sequential navigation (L-)subpatterns, i.e., frequent sequences of (L-)popular categories

Identifying interest groups

- Users sharing similar navigation behaviour and search interests
  - Searching for similar information in a similar way

Identifying interest groups

- Users sharing similar navigation behaviour and search interests
  - Searching for similar information in a similar way
- Interest groups construction
  - Exploit K-means clustering algorithm
  - Navigation patterns similarity
    - Ratio of the number of common categories (all their occurrences) to the total number of distinct categories
    - Example: navigation patterns
      $P_1 = \{$`Top,Arts,Classical_studies,Epigraphy,Latin,Epigraphy,Latin`$\}$ and
      $P_2 = \{$`Top,Arts,Classical_studies,Rome,Latin`$\}$
      4 common categories: `Top` ($\times 2$), `Arts` ($\times 2$), `Classical_Studies` ($\times 2$), `Latin` ($\times 3$)
      $S = 9/12 = 0.75$

Identifying interest groups

- Users sharing similar navigation behaviour and search interests
  - Searching for similar information in a similar way
- Interest groups construction
  - Exploit K-means clustering algorithm
  - Navigation patterns similarity
    - Ratio of the number of common categories (all their occurrences) to the total number of distinct categories
    - Example: navigation patterns
      $P_1 = \{$`Top,Arts,Classical_studies,Epigraphy,Latin,Epigraphy,Latin`$\}$ and
      $P_2 = \{$`Top,Arts,Classical_studies,Rome,Latin`$\}$
      4 common categories: `Top` ($\times 2$), `Arts` ($\times 2$), `Classical_Studies` ($\times 2$), `Latin` ($\times 3$)
      $S = 9/12 = 0.75$
  - Interest group = users whose navigation patterns in the same cluster
  - Each navigation pattern belongs to one cluster
  - User may belong to more than one interest groups

Identifying interest groups (cont'd)

Example

navigation patterns
{Top,Arts,Photography,Arts,Music,Dance}
{Top,Arts,Photography,Arts,Music,DJs}
{Top,Health,Medicine,Informatics,Journals_and_Publications}
{Top,Arts,Dance,Tango}
{Top,Computers,Information_Technology,Conferences}
{Top,Computers,Computer_Science,Publications,Bibliographies}

Construct 4 interest groups (clusters)

1. {Top,Arts,Photography,Arts,Music,Arts,Dance} and {Top,Arts,Dance,Tango}
2. {Top,Arts,Photography,Arts,Music,DJs}
3. {Top,Health,Medicine,Informatics,Journals_and_Publications}
4. {Top,Computers,Information_Technology,Conferences} and {Top,Computers,Computer_Science,Publications,Bibliographies}

Identifying indecisive users

Indecisive user

- Many "back and forth" visits to same categories
  - e.g. {`rock,80s,rock,80s,rock,60s,rock,60s`}
- This is due to:
  - Not knowing exactly what to search for in advance
  - Organization of categories different from user's intuitive categorization
  - Poor organization of topic sub-directories, or inconsistent category labels

Identifying indecisive users

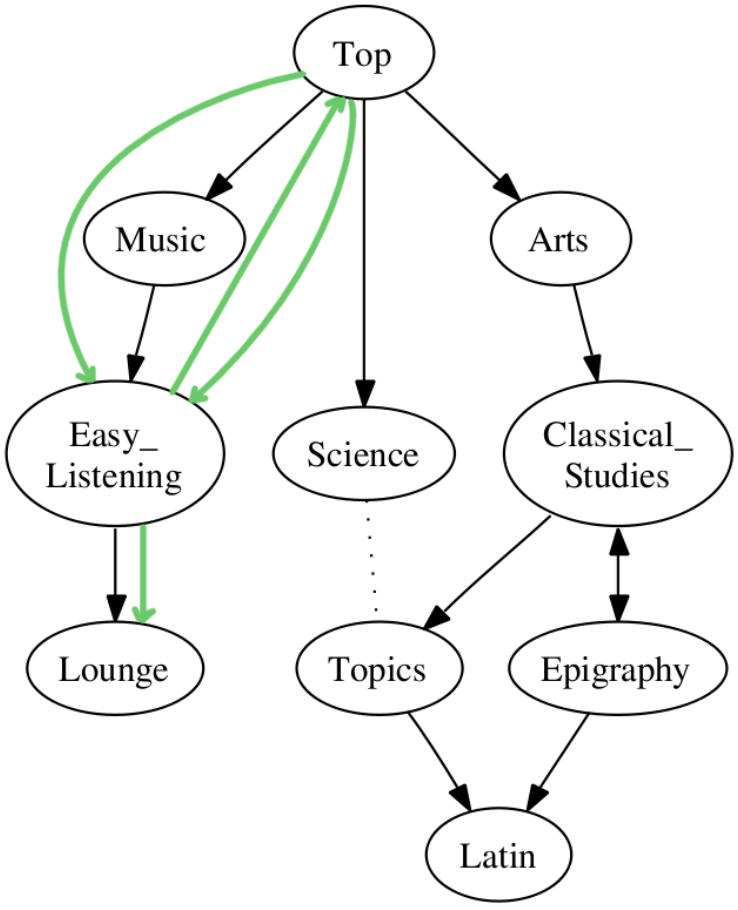
Indecisive user

- Many "back and forth" visits to same categories
  - e.g. {`rock,80s,rock,80s,rock,60s,rock,60s`}
- This is due to:
  - Not knowing exactly what to search for in advance
  - Organization of categories different from user's intuitive categorization
  - Poor organization of topic sub-directories, or inconsistent category labels

B&F actions/chains

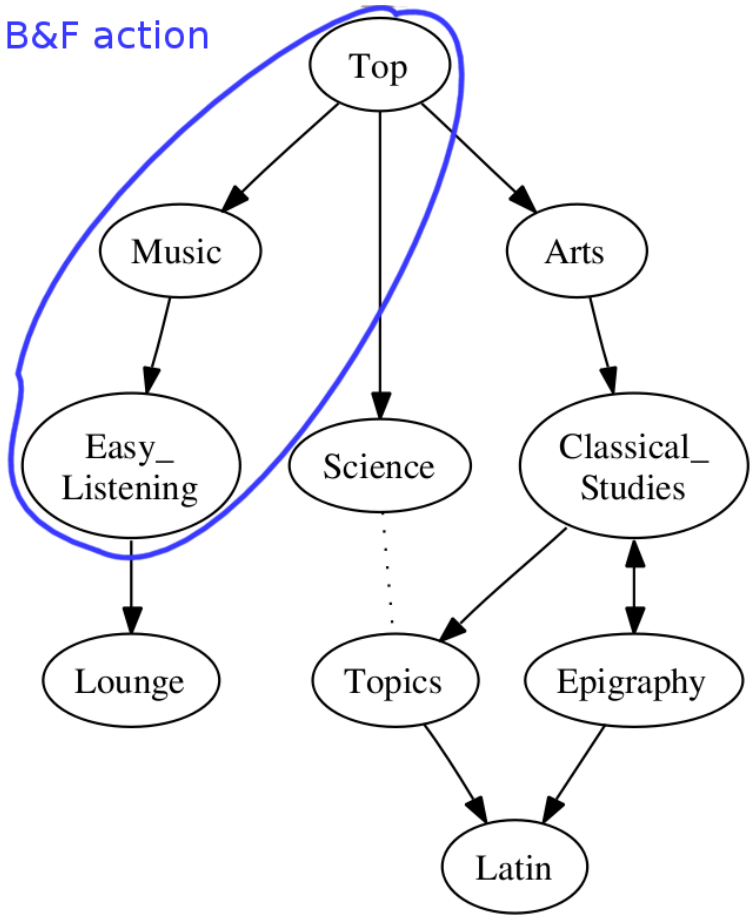
- Record B&F actions/chains to detect indecisive users
- For each navigation pattern check:
  - If exists sequence of categories $\{N_1, N_2, ..., N_k\}$ appearing twice
  - If between two occurrences, exists backwards action $\{N_{k-1}, ..., N_2\}$
  - B&F action $= \{N_1, N_2, ..., N_k\}$
  - B&F chain $= \{N_1, N_2, ..., N_k, N_{k-1}, ..., N_2, N_1, N_2, ..., N_k\}$

Identifying indecisive users (cont'd)



- Navigation pattern:
  {Top,Music,Easy_Listening,Music,Top,Music,Easy_Listening,Lounge}

Identifying indecisive users (cont'd)



- Navigation pattern:
  {Top,Music,Easy_Listening,Music,Top,Music,Easy_Listening,Lounge}
- B&F chain: {Top,Music,Easy_Listening,Music,Top,Music,Easy_Listening}

Mining (L-)popular categories & sequential navigation (L-)subpatterns

Two types of popular categories

- Popular: topics of great interest (i.e., frequently visited)
- L-popular: contain popular (i.e., frequently selected) resources
- Note that L-popular categories are not necessarily popular and vice versa

Mining (L-)popular categories & sequential navigation (L-)subpatterns

Two types of popular categories

- Popular: topics of great interest (i.e., frequently visited)
- L-popular: contain popular (i.e., frequently selected) resources
- Note that L-popular categories are not necessarily popular and vice versa

Sequential navigation (L-)subpatterns

- Frequent sequences of (L-)popular categories (i.e., frequent transitions (not necessarily contiguous) among (L-)popular categories)
- Not interested in identifying association rules
  - Because of the inherent order introduced by hierarchical organization of categories

Mining (L-)popular categories & sequential navigation (L-)subpatterns

Two types of popular categories

- Popular: topics of great interest (i.e., frequently visited)
- L-popular: contain popular (i.e., frequently selected) resources
- Note that L-popular categories are not necessarily popular and vice versa

Sequential navigation (L-)subpatterns

- Frequent sequences of (L-)popular categories (i.e., frequent transitions (not necessarily contiguous) among (L-)popular categories)
- Not interested in identifying association rules
  - Because of the inherent order introduced by hierarchical organization of categories

Identifying sequential navigation (L-)subpatterns

- Trie-based implementation [Bodon05] of Apriori [AS94] for mining frequent itemsequences
- Support: probability of visiting categories in the order specified in (L-)subpattern

Outline

Overview of personalization tasks

- Creation of shortcuts $A \longrightarrow B$, i.e. direct link from A to B
  - Alternative ways of navigating directory
  - Help users to easily reach topics tailored to their needs, while bypass others
  - Directed edge from A to B in the directory graph
- Two ways of creating shortcuts

Overview of personalization tasks

- Creation of shortcuts $A \longrightarrow B$, i.e. direct link from A to B
  - Alternative ways of navigating directory
  - Help users to easily reach topics tailored to their needs, while bypass others
  - Directed edge from A to B in the directory graph
- Two ways of creating shortcuts
  - Offline
    - Based on identifying frequent B&F chains and frequent sequential navigation (L-)subpatterns
    - Consider navigation patterns of each interest group
    - For each interest group, create static shortcuts
    - Present static shortcuts to all members of each group

Overview of personalization tasks

- Creation of shortcuts $A \longrightarrow B$, i.e. direct link from A to B
  - Alternative ways of navigating directory
  - Help users to easily reach topics tailored to their needs, while bypass others
  - Directed edge from A to B in the directory graph
- Two ways of creating shortcuts
  - Offline
    - Based on identifying frequent B&F chains and frequent sequential navigation (L-)subpatterns
    - Consider navigation patterns of each interest group
    - For each interest group, create static shortcuts
    - Present static shortcuts to all members of each group
  - Online
    - Based on identifying frequent sequential navigation (L-)subpatterns
    - Consider not only navigation patterns of "user's" interest groups
    - But also last categories visited in current user session
    - For each user, create dynamic shortcuts in real time
    - Present dynamic shortcuts to each individual user

Offline - Personalization based on frequent B&F chains

Shortcut creation

- Frequent B&F chains indicate difficulties for users in browsing
- This is due to:
  - Not knowing exactly what to search for in advance
  - Organization of categories different from user's intuitive categorization
  - Poor organization of topic sub-directories, or inconsistent category labels
- Bypass categories that confuse users or not tailored to their needs
- For each frequent B&F chain
  - A = first category of B&F chain
  - B = next category (in navigation pattern) after last one in B&F chain
  - Create shortcut A$\longrightarrow$B

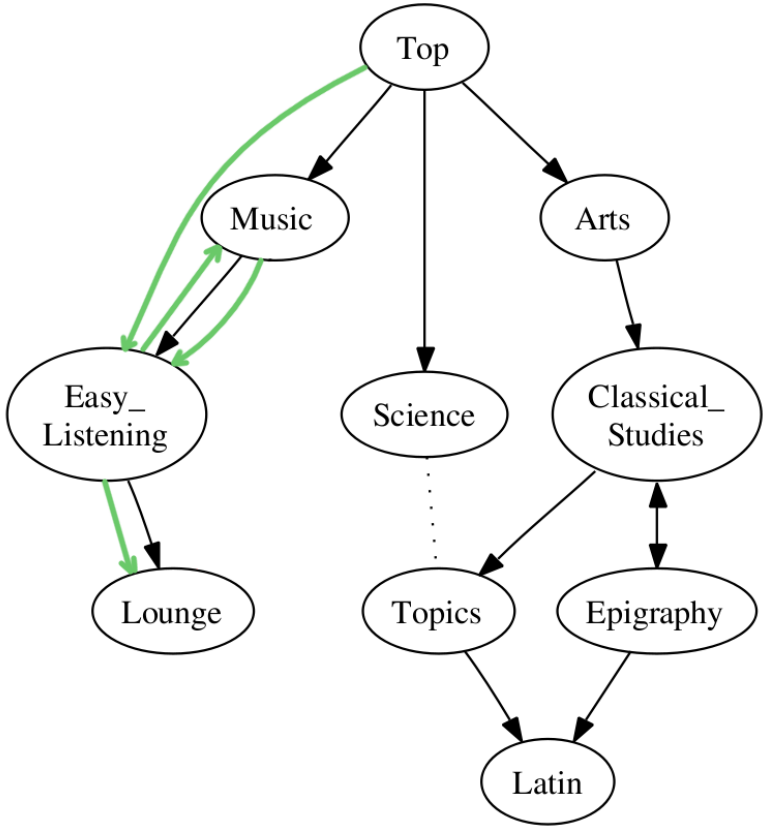
Offline - Personalization based on frequent B&F chains

Shortcut creation

- Frequent B&F chains indicate difficulties for users in browsing
- This is due to:
  - Not knowing exactly what to search for in advance
  - Organization of categories different from user's intuitive categorization
  - Poor organization of topic sub-directories, or inconsistent category labels
- Bypass categories that confuse users or not tailored to their needs
- For each frequent B&F chain
  - A = first category of B&F chain
  - B = next category (in navigation pattern) after last one in B&F chain
  - Create shortcut A⟶B

Example

- Navigation pattern: {Top,Music,Easy_Listening, Music,Easy_Listening,Lounge}



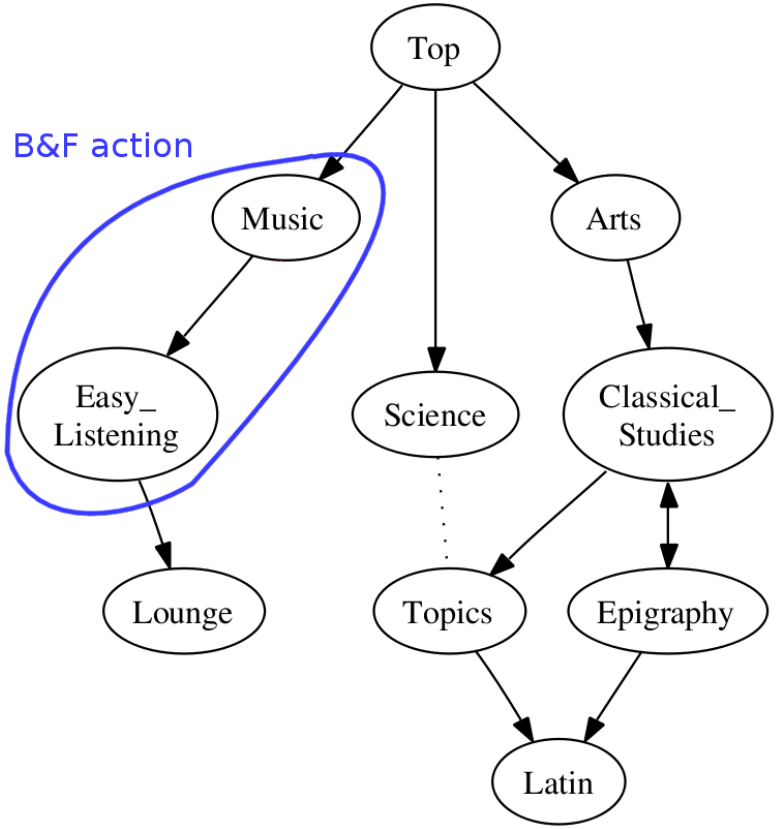
Offline - Personalization based on frequent B&F chains

Shortcut creation

- Frequent B&F chains indicate difficulties for users in browsing
- This is due to:
  - Not knowing exactly what to search for in advance
  - Organization of categories different from user's intuitive categorization
  - Poor organization of topic sub-directories, or inconsistent category labels
- Bypass categories that confuse users or not tailored to their needs
- For each frequent B&F chain
  - `A` = first category of B&F chain
  - `B` = next category (in navigation pattern) after last one in B&F chain
  - Create shortcut `A`⟶`B`

Example

- B&F chain: {`Music,Easy_Listening,Music, Easy_Listening`}



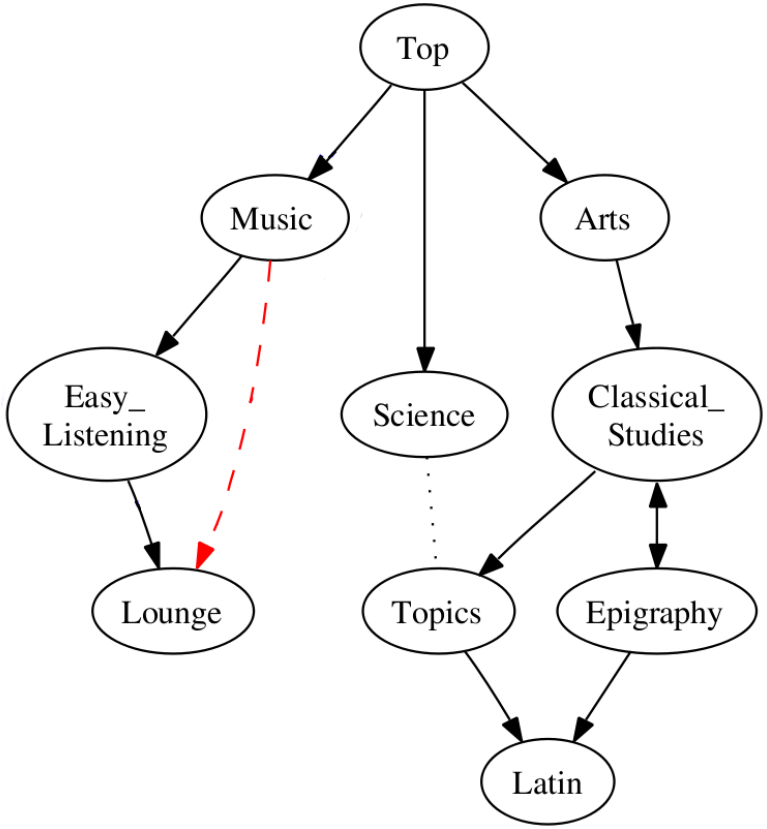
Offline - Personalization based on frequent B&F chains

Shortcut creation

- Frequent B&F chains indicate difficulties for users in browsing
- This is due to:
  - Not knowing exactly what to search for in advance
  - Organization of categories different from user's intuitive categorization
  - Poor organization of topic sub-directories, or inconsistent category labels
- Bypass categories that confuse users or not tailored to their needs
- For each frequent B&F chain
  - `A` = first category of B&F chain
  - `B` = next category (in navigation pattern) after last one in B&F chain
  - Create shortcut `A`⟶`B`

Example

- Assume B&F chain: {`Music`,`Easy_Listening`,`Music`, `Easy_Listening`} is frequent
- Create shortcut `Music`⟶`Lounge`



Offline - Personalization based on frequent sequential navigation (L-)subpatterns

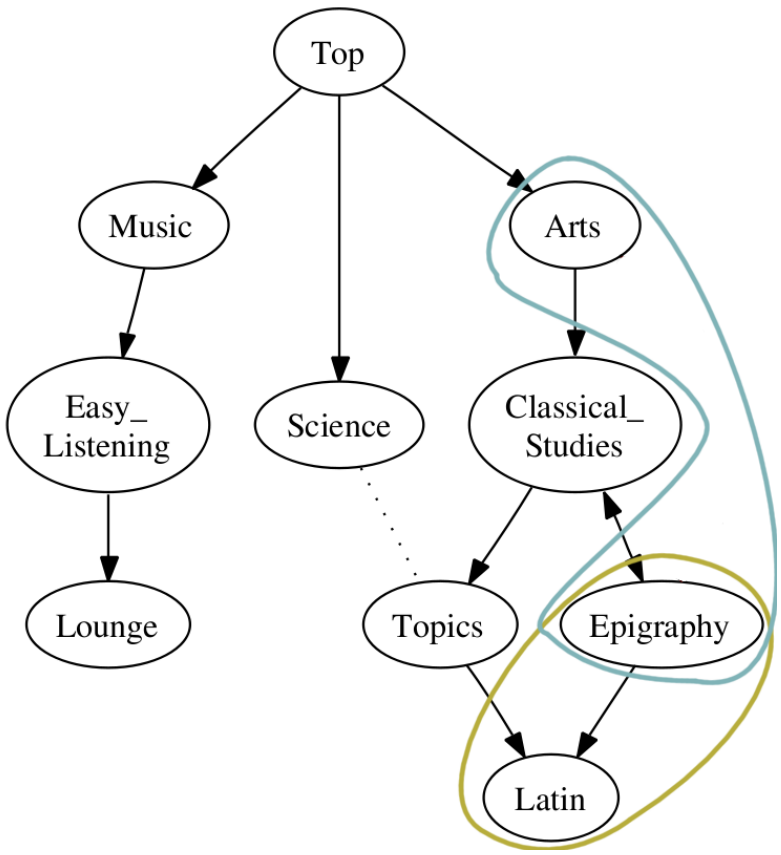
Shortcut creation

- Frequent sequential navigation (L-)subpatterns indicate popular transitions between (L-)popular categories
- Provide direct access to popular topics and resources
- For each interest group and a given support threshold
  - Identify 2-sequential navigation (L-)subpatterns $\{\mathtt{X},\mathtt{Y}\}$
  - Create shortcut $\mathtt{X} \longrightarrow \mathtt{Y}$

Offline - Personalization based on frequent sequential navigation (L-)subpatterns

Shortcut creation

- Frequent sequential navigation (L-)subpatterns indicate popular transitions between (L-)popular categories
- Provide direct access to popular topics and resources
- For each interest group and a given support threshold
  - Identify 2-sequential navigation (L-)subpatterns {X,Y}
  - Create shortcut X⟶Y



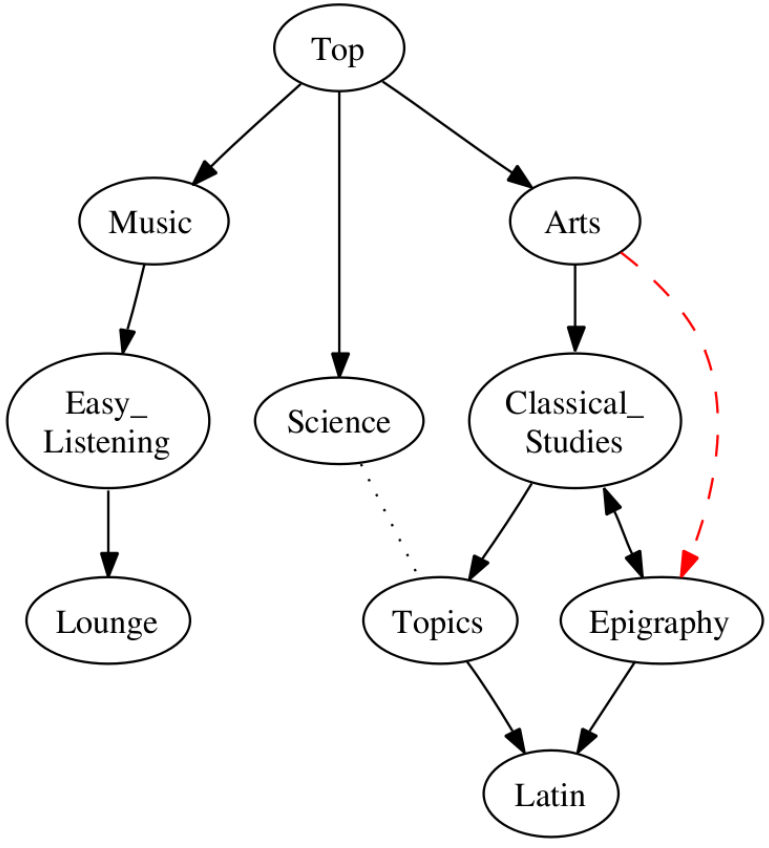
Example

- Frequent subpatterns: {Arts,Epigraphy} and {Epigraphy,Latin}
- Candidate shortcuts Arts⟶Epigraphy, Epigraphy⟶Latin

Offline - Personalization based on frequent sequential navigation (L-)subpatterns

Shortcut creation

- Frequent sequential navigation (L-)subpatterns indicate popular transitions between (L-)popular categories
- Provide direct access to popular topics and resources
- For each interest group and a given support threshold
  - Identify 2-sequential navigation (L-)subpatterns {X,Y}
  - Create shortcut X⟶Y



Example

- Frequent subpatterns: {Arts,Epigraphy} and {Epigraphy,Latin}
- Create shortcut Arts⟶Epigraphy

Online - Personalization based on frequent sequential navigation (L-)subpatterns

Active navigation window

- Retain two windows for each "user's" interest group
- Contains last $|w|$ (L-)popular categories visited

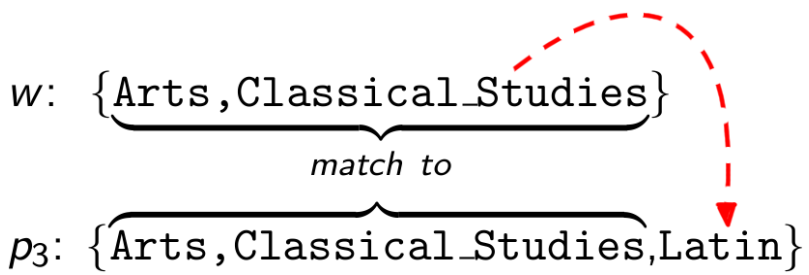
Shortcut creation

- Based on [MDL+02], but extended with multiple windows, interest groups
- For each interest group identify and store offline frequent sequential navigation (L-)subpatterns of size $|w| + 1$
- Match window with stored sequential navigation (L-)subpatterns
- For each matched frequent sequential navigation (L-)subpattern
  - `A` = last category of window
  - `B` = last category of (L-)subpattern
  - Create shortcut `A`$\longrightarrow$`B`, if its confidence is over given threshold
  - Confidence: conditional probability that user visits `B` provided that already visited all categories of window

Online - Personalization based on frequent sequential navigation (L-)subpatterns (cont'd)

Example

- Frequent sequential navigation subpatterns:
  p_1=$\{$`Arts`,`Classical_Studies`$\}$, support $\sigma(p_1) = 0.8$
  p_2=$\{$`Classical_Studies`,`Latin`$\}$, support $\sigma(p_2) = 0.7$
  p_3=$\{$`Arts`,`Classical_Studies`,`Latin`$\}$, support $\sigma(p_3) = 0.6$
- Assume $|w| = 2$, $w = \{$`Arts`,`Classical_Studies`$\}$
- Match w only to p_3 ($|p_3| = |w| + 1$, i.e., length acceptable)



- Shortcut `Classical_Studies`$\rightarrow$`Latin`
- $\alpha($`Classical_Studies`$\rightarrow$`Latin`$) = \frac{\sigma(p_3)}{\sigma(w)} = \frac{0.6}{0.8} = 0.75$

Outline

Evaluation method

Mining tasks - Precision and recall of interest groups

- 12 users
- 4 topics: video games, William Shakespeare, basketball, food and cooking
- 10 interest groups (clusters) created
- Interest groups precision and recall

Evaluation method

Mining tasks - Precision and recall of interest groups

- 12 users
- 4 topics: video games, William Shakespeare, basketball, food and cooking
- 10 interest groups (clusters) created
- Interest groups precision and recall

Offline personalization - Hit ration of static shortcuts

- Creation of static shortcuts
- Second period of user browsing
- Shortcut `A`$\longrightarrow$`B` hit ratio: number of times used to total times users moved from A to B

Evaluation method

Mining tasks - Precision and recall of interest groups

- 12 users
- 4 topics: video games, William Shakespeare, basketball, food and cooking
- 10 interest groups (clusters) created
- Interest groups precision and recall

Offline personalization - Hit ration of static shortcuts

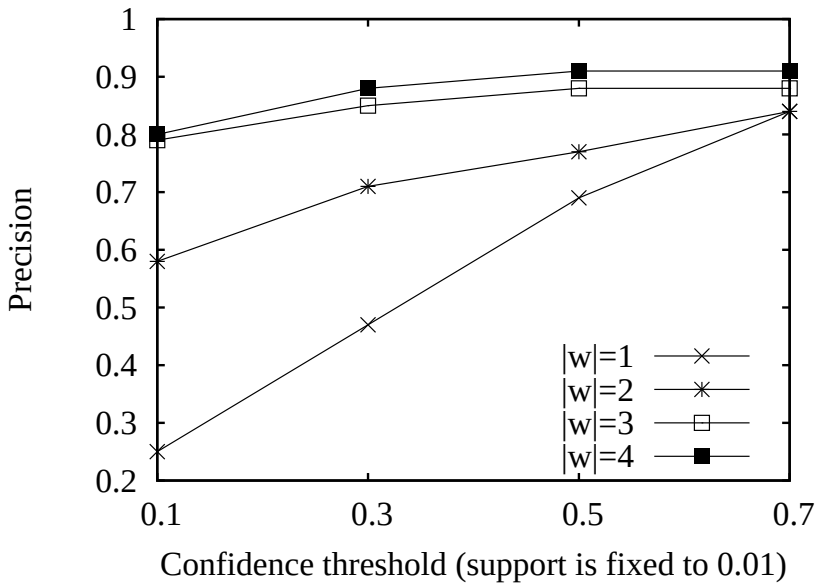
- Creation of static shortcuts
- Second period of user browsing
- Shortcut `A`$\longrightarrow$`B` hit ratio: number of times used to total times users moved from *A* to *B*

Online personalization - Precision of dynamic shortcuts

- Depth-first crawling at `Poetry`, `World_Literature` and `Drama` subtrees of `Top/Arts/Literature`
- Break navigation patterns
  - 70% generating dynamic shortcuts, 30% evaluation
- Shortcut `A`$\longrightarrow$`B` precision: number of categories *B* contained in 30% to total number of shortcuts

Online personalization - Precision of dynamic shortcuts (cont'd)

- Precision goes up as $|w|$ increases
  - Larger window provides a more representative part of user navigation behaviour
- Precision goes up as confidence threshold increases
  - Increased confidence for A⟶B means high probability that B in 30% part of navigation patterns
- Precision goes up as support threshold increases



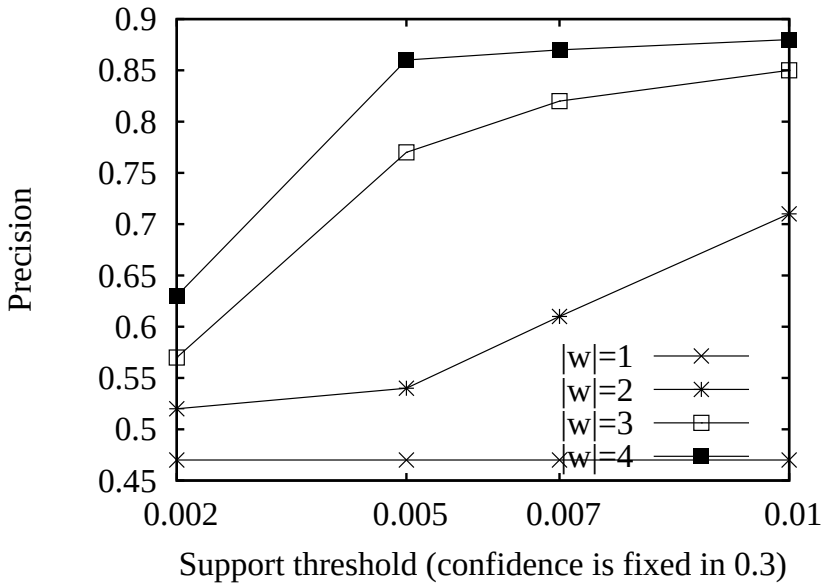


Figure: Precision of the personalization task varying the confidence/support threshold for several values of $|w|$.

Conclusion - Future work

Conclusion

- Methodology for personalizing topic directories according to users navigation behaviour
  - Set of mining tasks: interest groups, indecisive user behaviour, frequent navigation (L-)subpatterns
  - Set of personalization tasks: shortcuts creation
- Experiments for evaluating mining and personalization tasks

Future work

- Enhance personalization tasks
  - User-driven profiles
  - Semantically rich topic directories, e.g. *IS_A*, *PART_OF* relationships
- Extend evaluation of online personalization - study real user navigation patterns

Thank you

http://casablanca.dblab.ece.ntua.gr/p-miner

Related work

- Discovering sequences of visits
  - Datamining techniques
  - Probabilistic models
  - Most of them, do not perform personalization
  - The rest, do not distinguish between different users and groups of users
- Personalization in Digital Libraries and Web portals
  - The structure of these Web sites is similar to topic directories
  - Based on explicit user input
    - Provide simplified search functionalities and alerts
  - Based on implicit user input
    - They identify the preferences of each individual user
- Collaborative filtering-based methods
  - Also identify users with common interests and behaviour
  - Model user profiles as vectors
  - On the contrary, we use clustering to create interest groups
  - Also exploit sequential pattern mining to generate recommendations

System architecture