A Recommendation System for Software Function Discovery

Naoki Ohsugi

Software Engineering Laboratory,

Graduate School of Information Science,

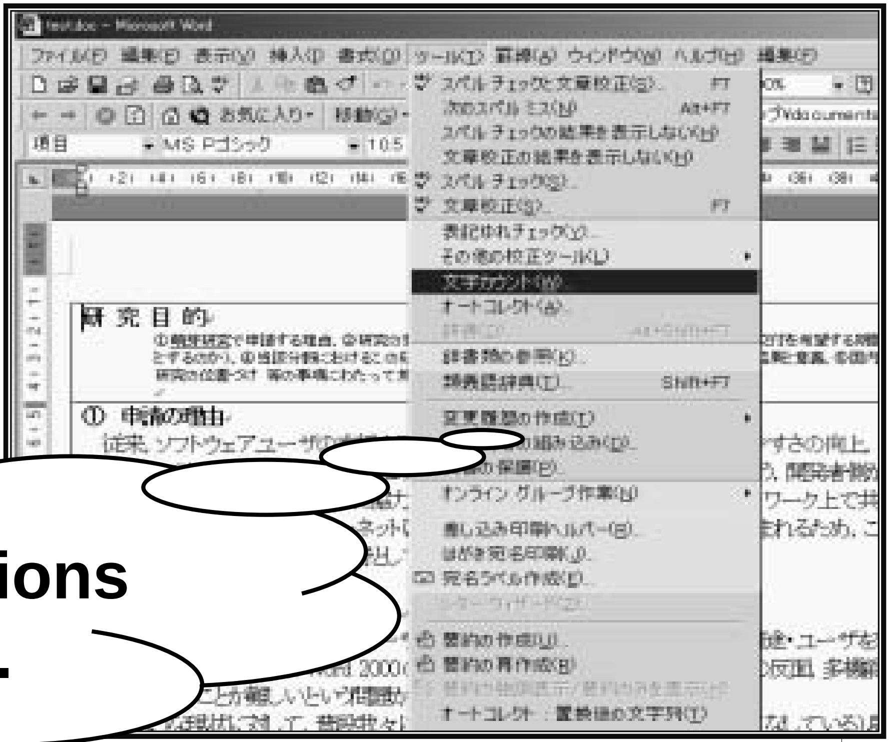
Nara Institute of Science and Technology

Tuesday 16 December, 2003.

International Workshop on Community-Driven Evolution of Knowledge Artifacts

Growth of Software Functions

- Application software is getting more complicated and providing more functions.
  - Total number of menu items (Microsoft Office)
    - **Word 2000: 660**
    - **Word 2002: 772**
    - **Excel 2000: 705**
    - **Excel 2002: 792**
    - **PowerPoint 2000: 565**
    - **PowerPoint 2002: 646**

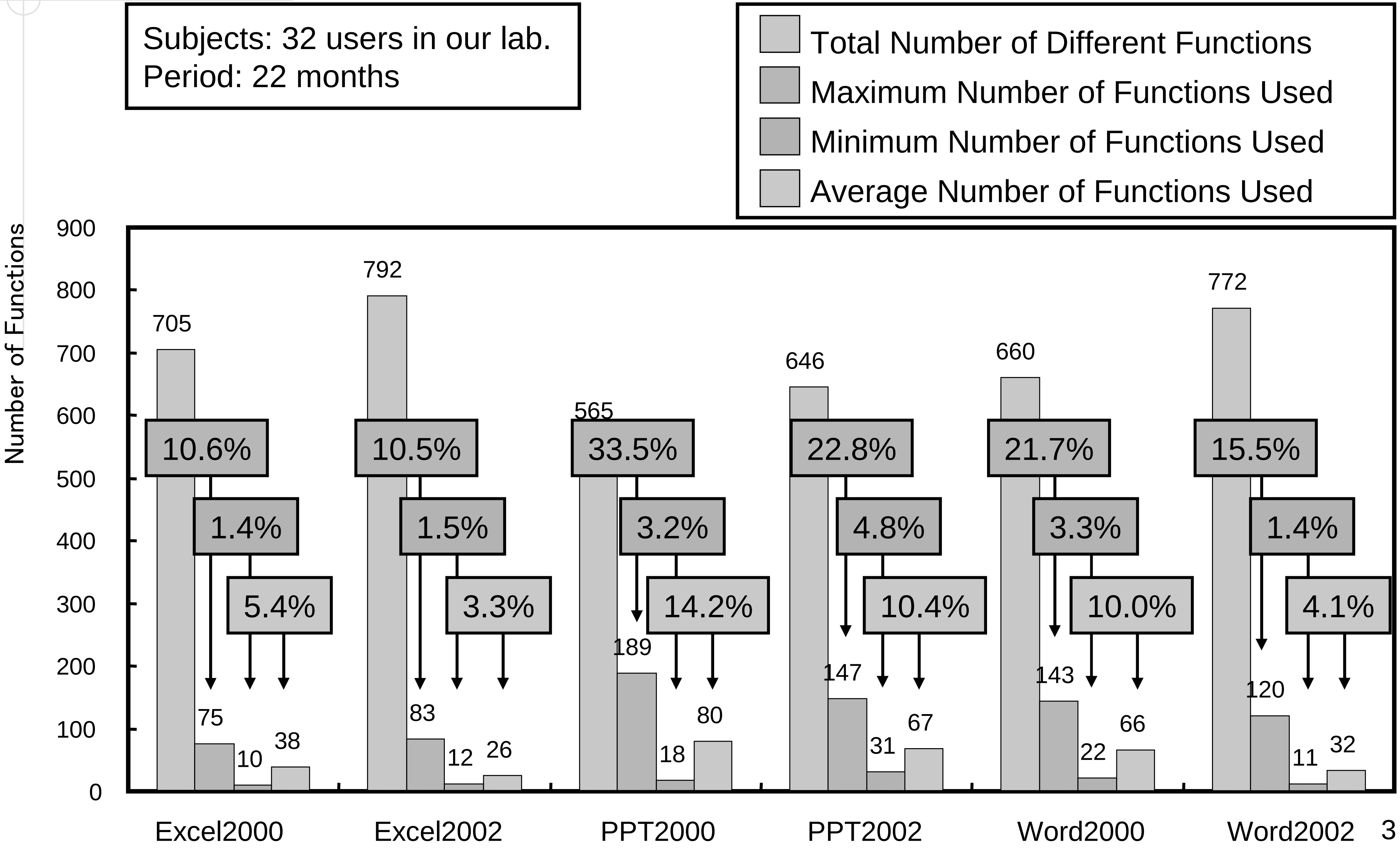

test.doc - Microsoft Word
ファイル(F) 編集(E) 表示(V) 挿入(I) 書式(O) ツール(T) 罫線(A) ウィンドウ(W) ヘルプ(H)
スペル チェックと文章校正(S)... F7
次のスペル ミス(N) Alt+F7
スペル チェックの結果を表示しない(K)
文章校正の結果を表示しない(H)
スペル チェック(S)...
文章校正(G)... F7
表記ゆれチェック(Y)...
その他の校正ツール(L)
文字カウント(W)...
オートコレクト(A)...
辞書類の参照(K)...
類義語辞典(T)... Shift+F7
変更履歴の作成(T)
オンライン グループ作業(N)
差し込み印刷ヘルパー(R)...
はがき宛名印刷(J)...
宛名ラベル作成(L)...
要約の作成(U)...
要約の再作成(B)
オートコレクト：置換後の文字列(T)
研 究 目 的
① 申請の理由

Users can’t find useful functions from too many functions.

Screenshot of MS-Word 2002

Users Could Not Find Some Useful Functions!

Subjects: 32 users in our lab.
Period: 22 months

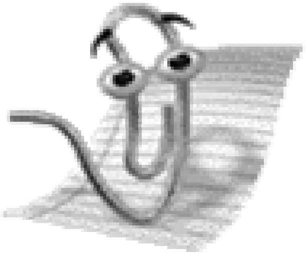


A Recommendation System for Software Function Discovery

- The system recommends individual users a set of candidate functions, which may be useful.
- Our solution is a Collaborative Filtering approach.

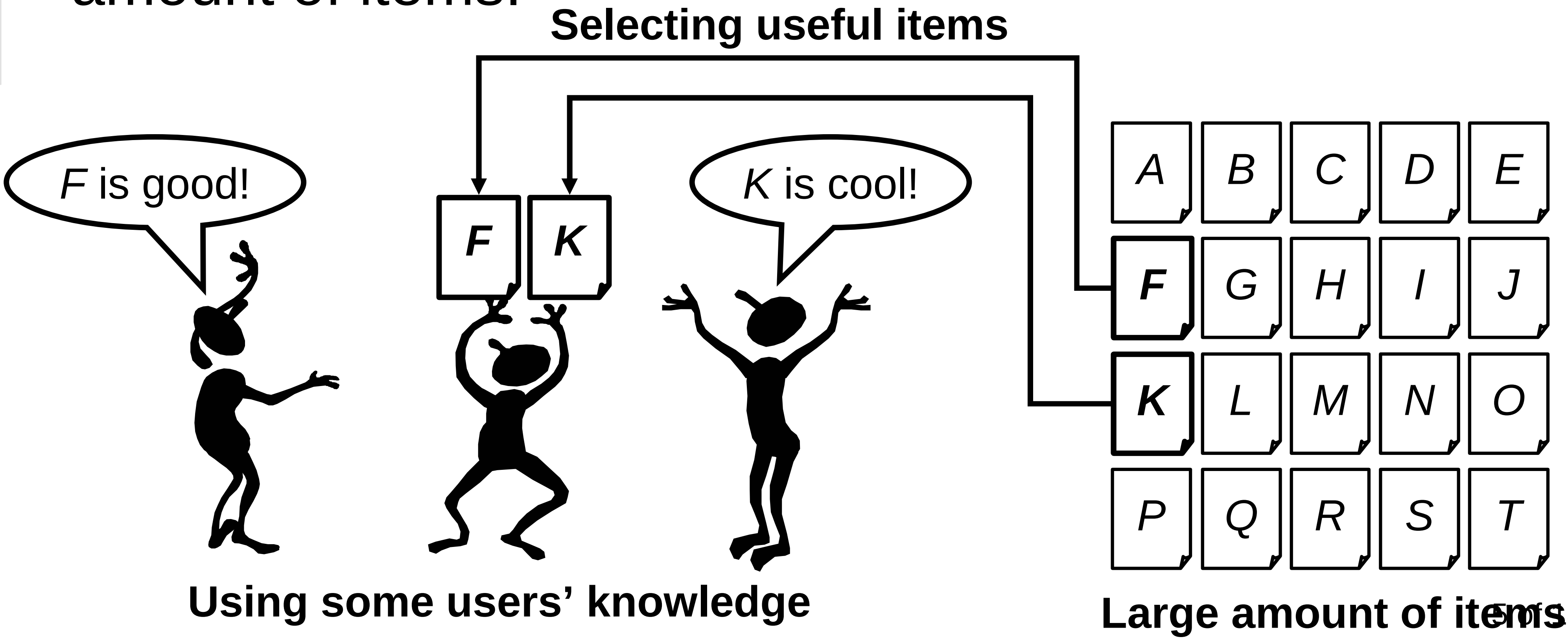
Here’s my recommendation:

Tools → Word Count…	21 pts
Insert → Date Time…	20 pts
Tools → Thesaurus…	18 pts
Insert → Footnote…	18 pts
Tools → Spelling…	17 pts



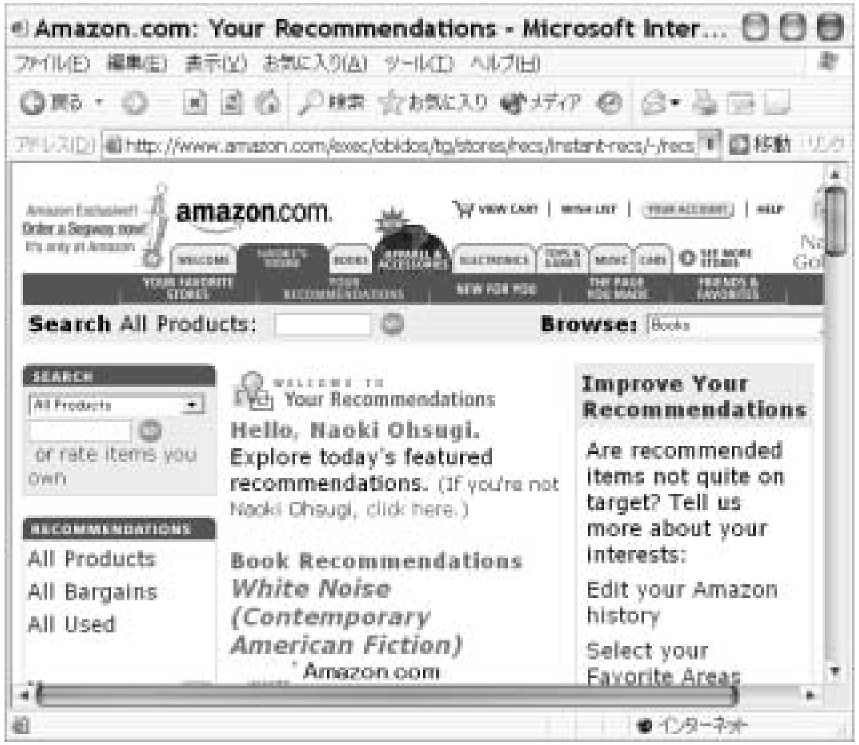
What is Collaborative Filtering (CF)?

- **“Collaborative**” means using some users’ knowledge for filtering.
- **“Filtering”** means selecting useful items from large amount of items.



Voting-based Recommendation Systems with CF

- The systems collect *explicit* votes as users' knowledge.



Amazon.com: Your Recommendations - Microsoft Inter...

http://www.amazon.com/exec/obidos/tg/stores/recs/instant-recs/-/recs

Search All Products:

Browse: Books

WELCOME TO Your Recommendations

Hello, Naoki Ohsugi. Explore today's featured recommendations. (If you're not Naoki Ohsugi, click here.)

Book Recommendations

White Noise (Contemporary American Fiction)

Amazon.com

Improve Your Recommendations

Are recommended items not quite on target? Tell us more about your interests:

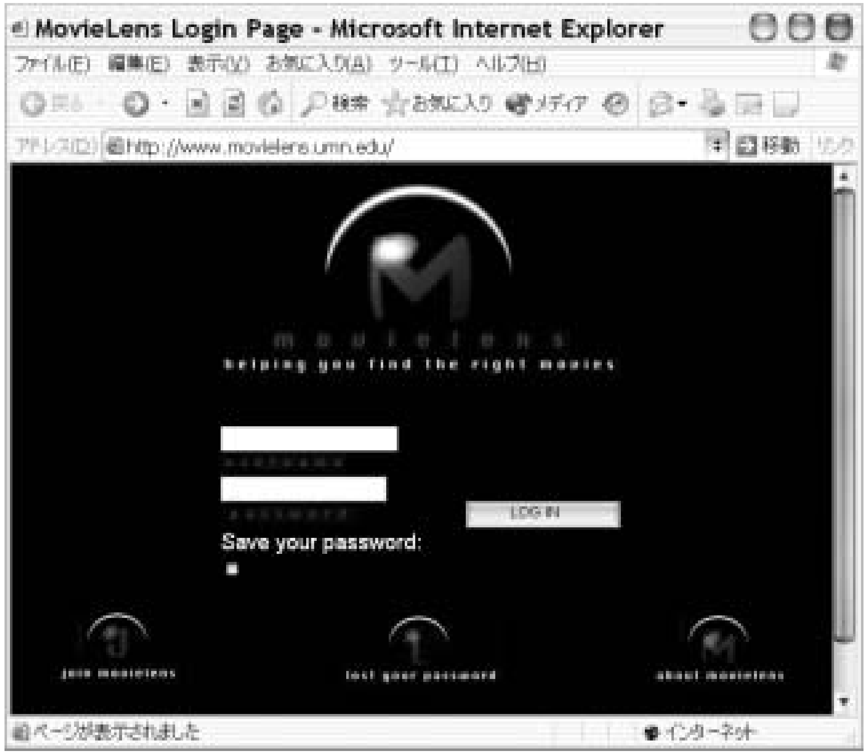
Edit your Amazon history

Select your Favorite Areas

Amazon.com

(Book recommendation system)

http://www.amazon.com



MovieLens Login Page - Microsoft Internet Explorer

http://www.movielens.umn.edu/

movielens

helping you find the right movies

Save your password: □

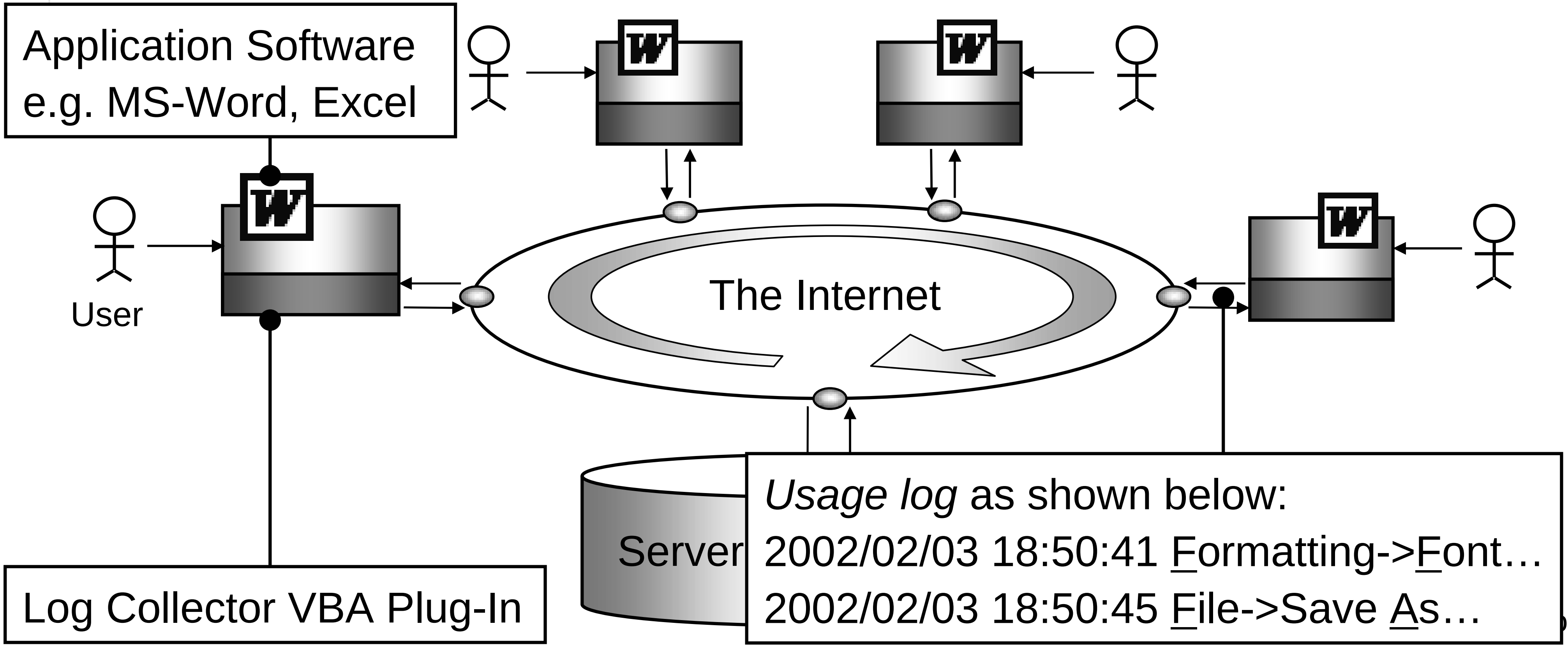
MovieLens

(Movie recommendation system)

http://www.movielens.umn.edu

Logging Usage as Users' Knowledge

- The proposed system automatically collects the records of executed functions (*Usage logs*) as users' knowledge.
- *Usage logs* are collected from some users via the Internet.

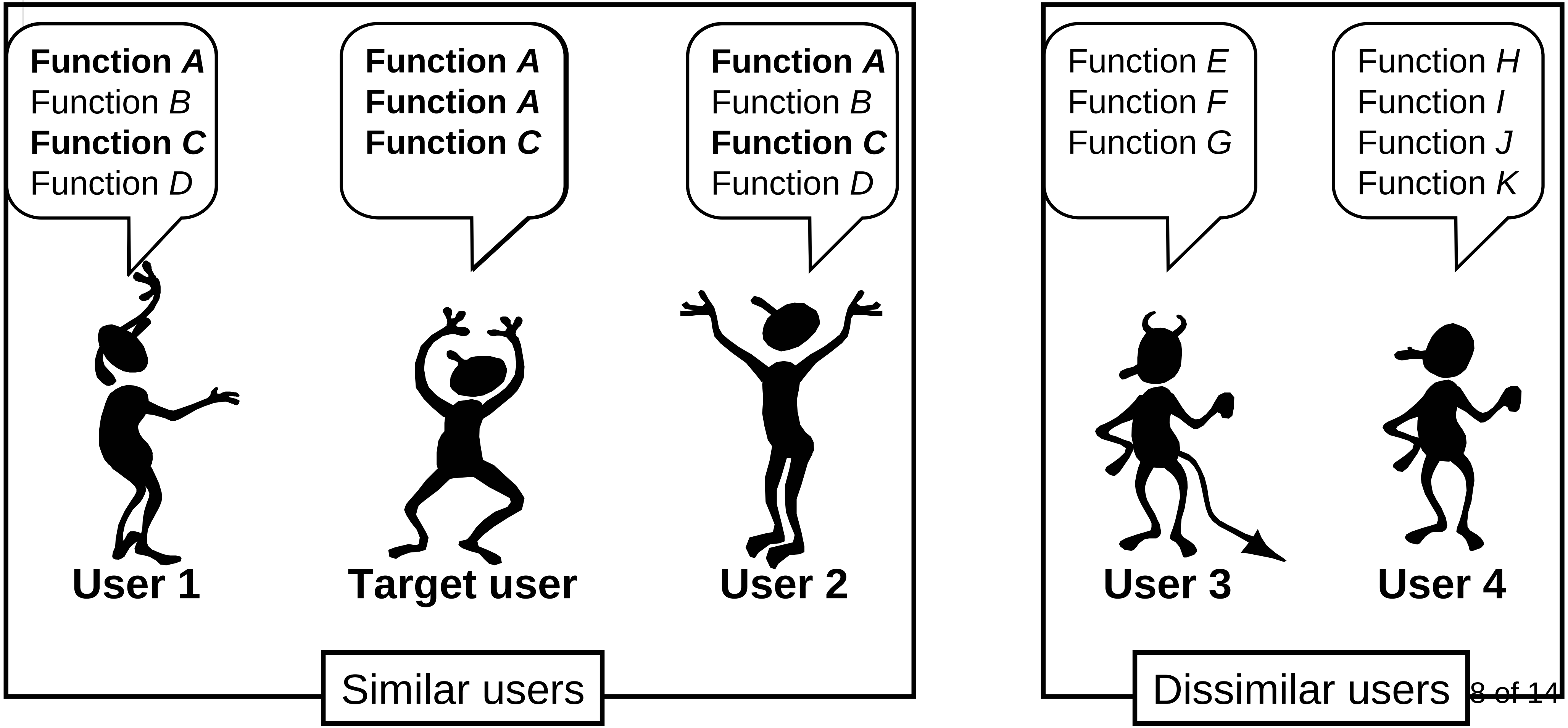


Usage log as shown below:

```
2002/02/03 18:50:41 Formatting->Font…
2002/02/03 18:50:45 File->Save As…
```

Step1: Computing Similarities

- Computing similarities between the target user and the other users



Function *A*
Function *B*
Function *C*
Function *D*

User 1

Function *A*
Function *A*
Function *C*

Target user

Function *A*
Function *B*
Function *C*
Function *D*

User 2

Similar users

Function *E*
Function *F*
Function *G*

User 3

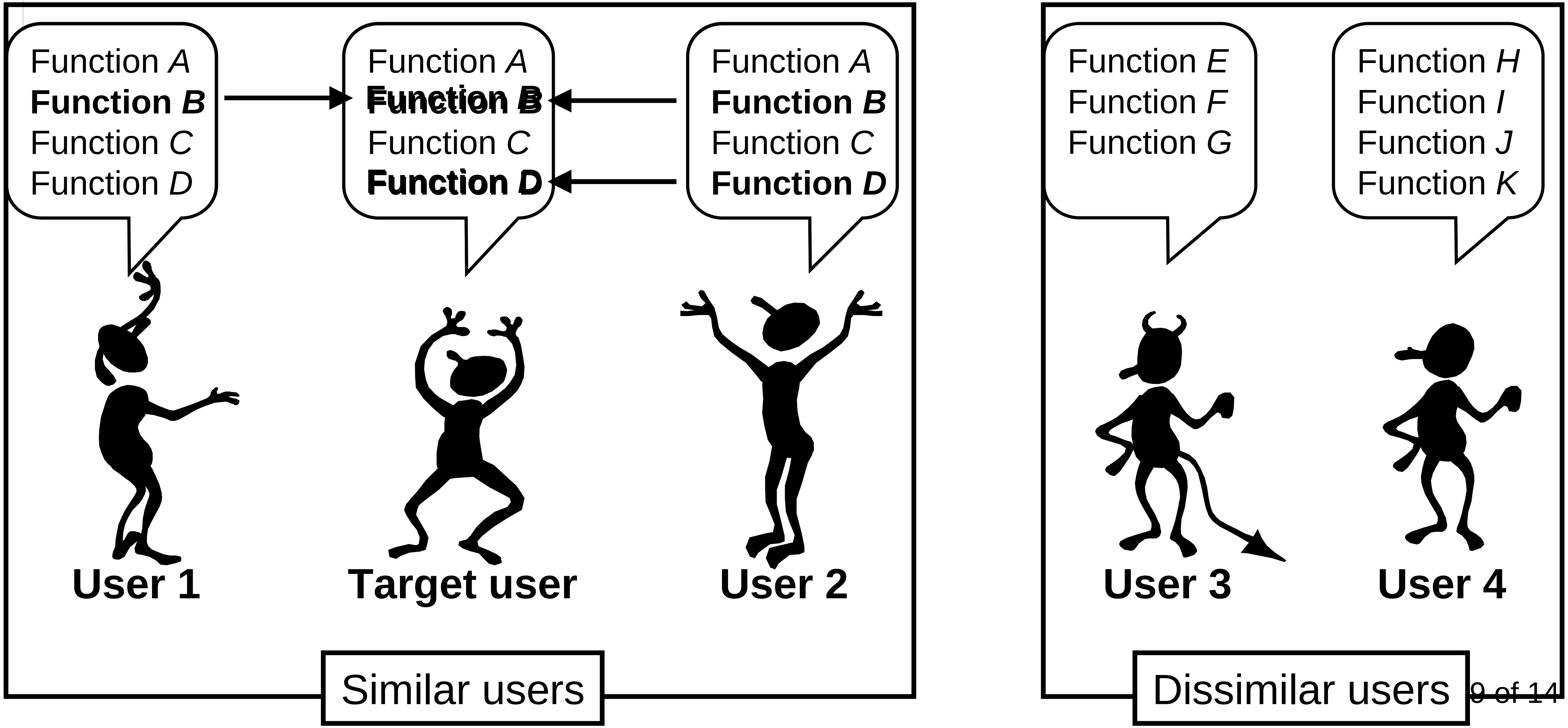
Function *H*
Function *I*
Function *J*
Function *K*

User 4

Dissimilar users

Step 2: Delivering Knowledge

- Delivering the useful functions candidate, which were frequently used by the similar users'.



Function *A*
Function *B*
Function *C*
Function *D*

Function *A*
Function *B*
Function *C*
Function *D*

Function *A*
Function *B*
Function *C*
Function *D*

Function *E*
Function *F*
Function *G*

Function *H*
Function *I*
Function *J*
Function *K*

User 1
Target user
User 2
User 3
User 4

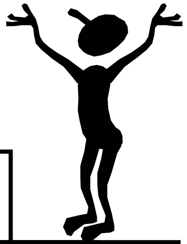
Similar users

Dissimilar users

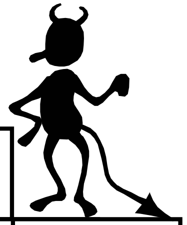
Conventional Similarity Calculation

- Calculating Similarities by Correlation Coefficient
  - The dominant frequencies (e.g., “Undo” or “Save”) over-affect similarity computations.

Target user		
1	Undo	60%
2	Save	20%
3	Redo	10%
4	Copy	4%
5	Paste	3%
6	Cut	2%
7	Clear	1%



User 2		
1	Save	55%
2	Undo	25%
3	Redo	10%
4	Copy	4%
5	Paste	3%
6	Cut	2%
7	Clear	1%



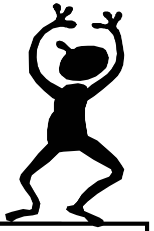
User 3		
1	Undo	60%
2	Save	20%
3	Clear	6%
4	Cut	5%
5	Copy	4%
6	Paste	3%
7	Redo	2%

Correlation based similarity	+0.41	+0.97

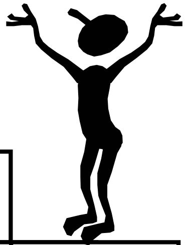
(Range of value [-1.00, +1.00])

Better Similarity Calculation

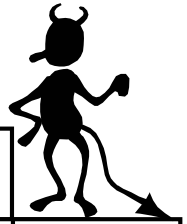
- Calculating Similarities by Rank Correlation
  - The dominant frequencies ("Undo" & "Save") do not affect similarity computations.



Target user		
1	Undo	60%
2	Save	20%
3	Redo	10%
4	Copy	4%
5	Paste	3%
6	Cut	2%
7	Clear	1%



User 2		
1	Save	55%
2	Undo	25%
3	Redo	10%
4	Copy	4%
5	Paste	3%
6	Cut	2%
7	Clear	1%

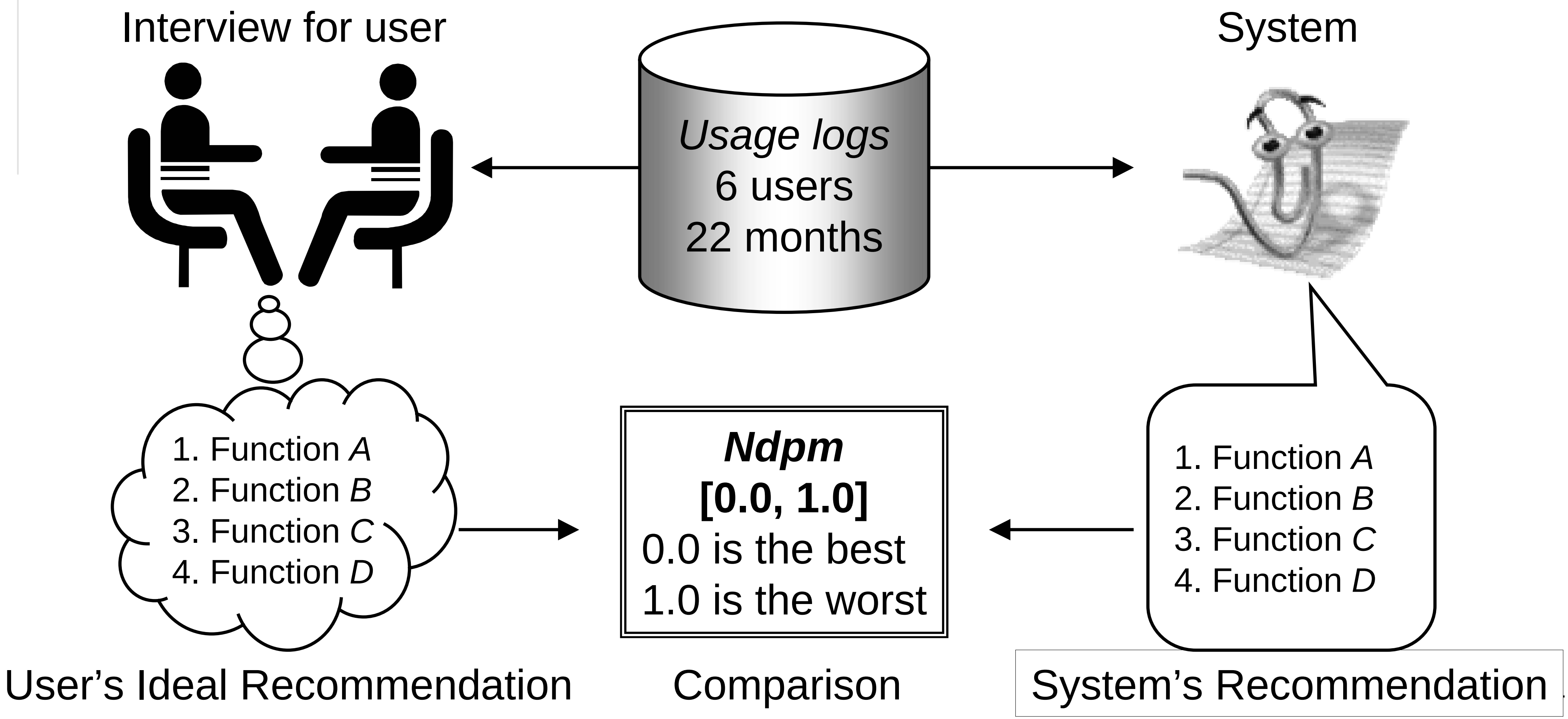


User 3		
1	Undo	60%
2	Save	20%
3	Clear	6%
4	Cut	5%
5	Copy	4%
6	Paste	3%
7	Redo	2%

	User 2	User 3
Correlation based similarity	+0.41	+0.97
Rank correlation based similarity	+0.90	+0.05

Evaluating Accuracy of Recommendation

- Yao's *ndpm* measure
  - * Y.Y. Yao, "Measuring Retrieval Effectiveness Based on User Preference of Documents", *J. of American Society for Information Science*, 46, 2, 1995, pp.133-145.



Interview for user

Usage logs
6 users
22 months

System

1. Function *A*
2. Function *B*
3. Function *C*
4. Function *D*

Ndpm
[0.0, 1.0]
0.0 is the best
1.0 is the worst

1. Function *A*
2. Function *B*
3. Function *C*
4. Function *D*

User's Ideal Recommendation

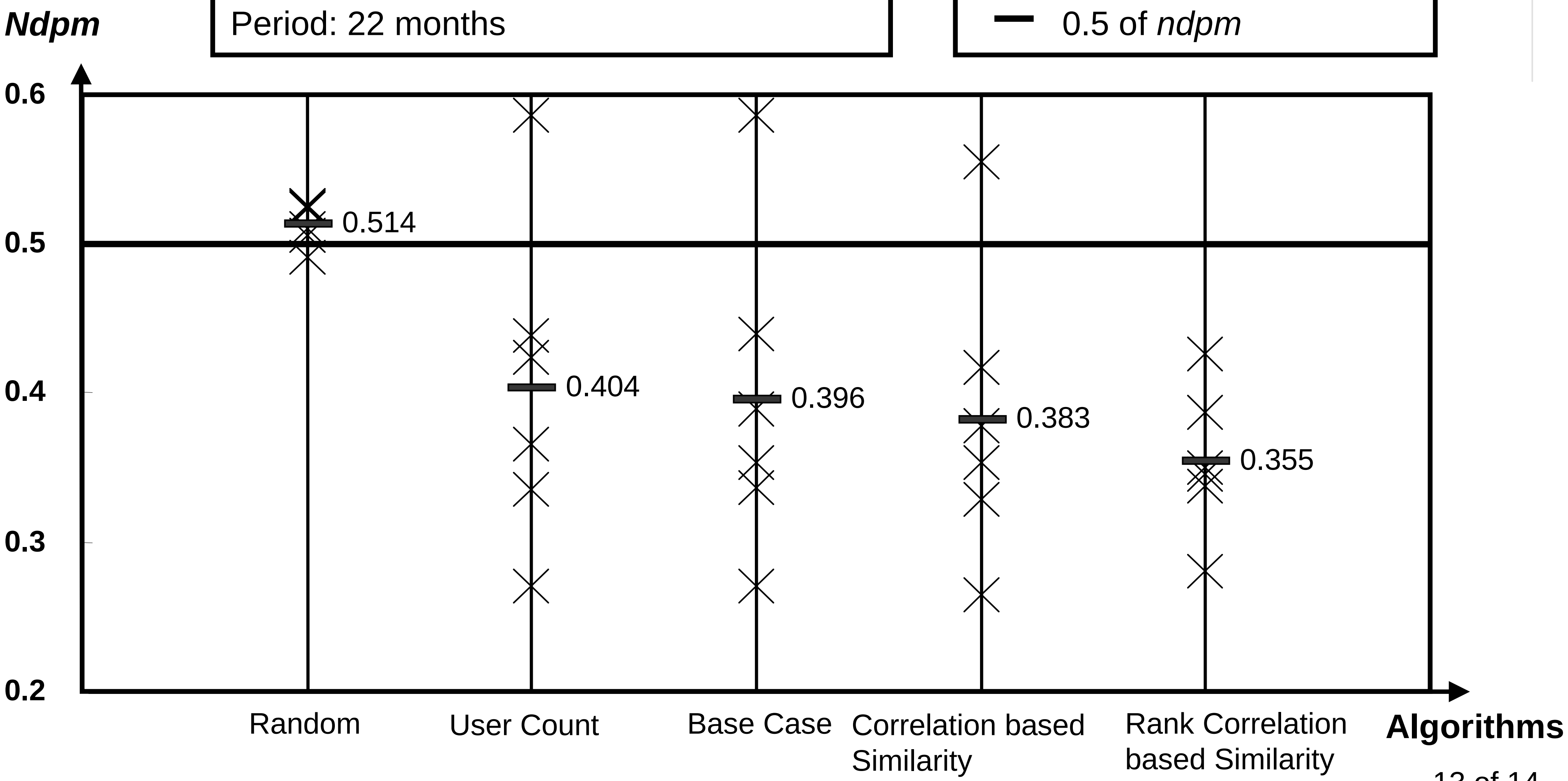
Comparison

System's Recommendation

Experimental Result

Collected usage logs of Ms-Word 2000
Subjects: 6 users in our lab.
Period: 22 months

× Each user's *ndpm*
— Average of *ndpm*
— 0.5 of *ndpm*



Conclusion

- I proposed a recommendation system to help users discover useful functions.

- I evaluated the accuracy of recommendation.
  - The result suggested the proposed system has a potential to provide useful recommendation for software function discovery.