User Modeling and Recommender Systems

Test Exam

(Dr. W.Wörndl)

Name: _____

Matrikelnr.: _____

Date: ...

This exam consists of 6 tests with a total maximum of 40 points. The duration of the exam is 60 minutes. Lecture material or other auxiliary material is not allowed during the exam. Please write your name on all 8 pages of the exam.
Complete sentences are not required, you can answer by using (meaningful) catchwords and/or labeled drawings. If the provided space is not sufficient, you can use the back side of the pages. You can answer in either German or English. /
Sie können Ihre Antwort in deutscher oder englischer Sprache formulieren.

Grading:

1.0	*38-40*	*2.0*	*29-31*	*3.0*	*23-24*	*4.0*	*17-18*
1.3	*35-37*	*2.3*	*27-28*	*3.3*	*21-22*	*4.3*	*10-16*
1.7	*32-34*	*2.7*	*25-26*	*3.7*	*19-20*	*4.7*	*5-9*

Points between the grades will lead to the better grade (e.g.: 34.5 P. → 1,3).

T1	T2	T3	T4	T5	T6	Σ *Erstkorrektur*
T1	T2	T3	T4	T5	T6	Σ *Zweitkorrektur*

Grade

Test 1: Definitions (7 Points)

a) Define the term *"context"* as introduced in this course and give a short example how context can be used to improve recommender systems. *(3 P.)*

b) What is a *unary* rating scale? *(2 P.)*

c) Briefly explain a method for *non-intrusive user identification*. *(2 P.)*

Test 2: (Web) Document Modeling (7 P.)

Document modeling aims at constructing an abstract representation of a document. The result is a vector of index terms with their relevance for discriminating the document using methods such as TF*IDF.

a) Briefly explain how to match a user query with the document vector using the *vector space model (VSM)*. *(4 P.)*

b) *"Tag-based Web-IR"* is one of the options to improve standard document modeling in the WWW. Briefly explain the main idea behind *Tag-based Web-IR* and outline how this approach can be integrated in the classic IR model. *(3 P.)*

Test 3: Hybrid Recommender Systems and Social Choice (7 P.)

a) Briefly explain the *Cascading* method of hybrid recommender systems. *(3 P.)*

b) Briefly explain the *"Borda Count"* strategy for social choice using a self-selected example with 2 users and 3 items. *(4 P.)*

Test 4: Recommender Systems (7 P.)

a) Briefly explain two disadvantages of collaborative filtering recommender systems. *(3 P.)*

b) Briefly explain the *„Bounded Greedy Selection“* algorithm to improve diversity in a case-based recommender system. *(4 P.)*

Test 5: Evaluation (4 P.)

a) Briefly explain how to evaluate a recommender system using *"k-fold cross-validation"*, given a dataset of user interactions (e.g. ratings). *(3 P.)*

b) What is the best metric if you want to measure how many of your recommendations were successful? (no explanation needed) *(1 P.)*

Test 6: Single Choice (8 P.)
Mark the correct answer to the following questions. Only 1 out of the 3 answers is correct. (No justification needed.)

a) What is the correct description of the metric "*confidence"* in association rule mining (for rules X → Y).

- ○ Confidence is the number of times Y appears in transactions divided by the total number of transactions
- ○ Confidence measures how often items in Y appear in transactions that contain X
- ○ Confidence is the fraction of transactions that contain both X and Y

b) Why is it not always preferable to just present the items with the highest predicted rating to the user?

- ○ Data sparseness of user-item matrix
- ○ Diversity of recommendations may be poor
- ○ High predicted rating correspond to low user acceptance

c) Which of the following is <u>not</u> a reasonable *convergence criteria* for *k-means clustering*?

- ○ Centroid partitions do not change
- ○ Object partitions do not change
- ○ Object – centroid distance is larger than 0

d) What is the correct assumption of *loose coupling* of context and user preferences?

- ○ Both user preferences and item consumption depend on the context
- ○ Item consumption doesn't depend on the context
- ○ User preferences don't depend on the context

e) What is the goal of the *matrix factorization* process?

- ○ Calculate estimations of unknown values of the rating matrix
- ○ Group a set of items into classes of similar items
- ○ Determine rules that will predict the occurrence of items from the occurrence of other items

f) Which type of recommender system is best suited for a *low risk and heterogeneous domain with implicit interaction style*?

- ○ Collaborative filtering
- ○ Content-based recommender system
- ○ Knowledge-based recommender system

g) Hiding or removing a link in adaptive hypermedia is reasonable for which type of link?

- ○ Contextual
- ○ Non-contextual
- ○ Intentional

g) Which of the following is <u>not</u> an action *mix stations* in a mix network perform to achieve unobservable and unlinkable message exchange?

- ○ Change order of message packets
- ○ Intersperse dummy messages
- ○ Modify the content of the message