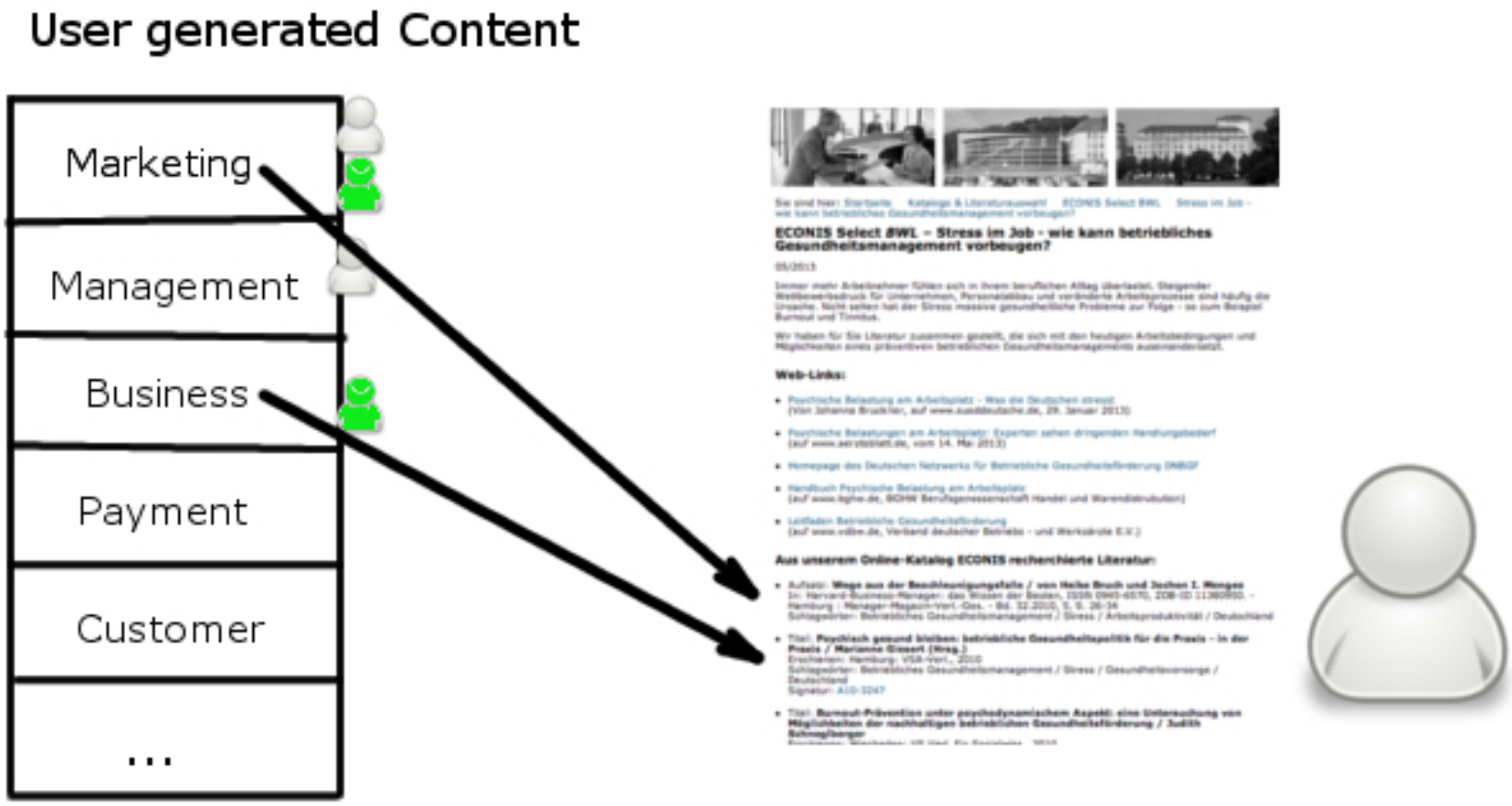
A Contribution to Rating and Recommendation Systems: Concepts, Development and Evaluation

Oliver Diestel

June 10, 2013

The Problem

The problem: Recommending user generated content.



Challenges

Verweildauer des Managements

0

Gibt es wissenschaftliche Studien (oder vergleichbares) zu der Verweildauer des Managements in einem Unternehmen? Oder anderesherum: gibt es Studien über den sog. management turnover (Fluktuation), also wie häufig das Management ausgewechselt wird?

asked **3 mins ago**
newTester
3 • 1

edit flag offensive close delete

post a comment

1 Answer Sort by » oldest newest most voted

0

Änderungen im Vorstand (bei AGs) oder der Geschäftsführung (bei GmbHs) müssen dem zuständigen Handelsregister beim Amtsgericht gemeldet werden. Inwieweit solche Meldungen statistisch erfasst werden, müssten Sie beim Statistischen Bundesamt

answered **2 mins ago**
test
46 • 2

updated **1 mins ago**

http://www.destatis.de/jetspeed/portal/cms/Sites/destatis/Internet/DE/Navigation/InfoService/InfoService.ps oder beim Registerportal des Bundes https://www.handelsregister.de/rp_web/contact.do erfragen. Auf den Webseiten der Institutionen finden sich keine Hinweise darauf.

Studien zum Thema finden Sie z.B. in unserer Online-Datenbank ECONIS, wenn Sie mit dem Titelstichwort ""management turnover"" (in Anführungszeichen gesetzt wird die komplette Phrase gesucht) suchen oder aber mit den sachlichen Schlagworten Fluktuation UND Deutschland und dann weiter mit dem Titelstichwort Management eingrenzen. Die Literaturverzeichnisse der gefundenen Werke geben evtl. weitere Hinweise auf Studien.

Ich hoffe, dass Ihnen diese Angaben weiterhelfen. Bei Rückfragen zu dieser Mail oder anderen Fragen

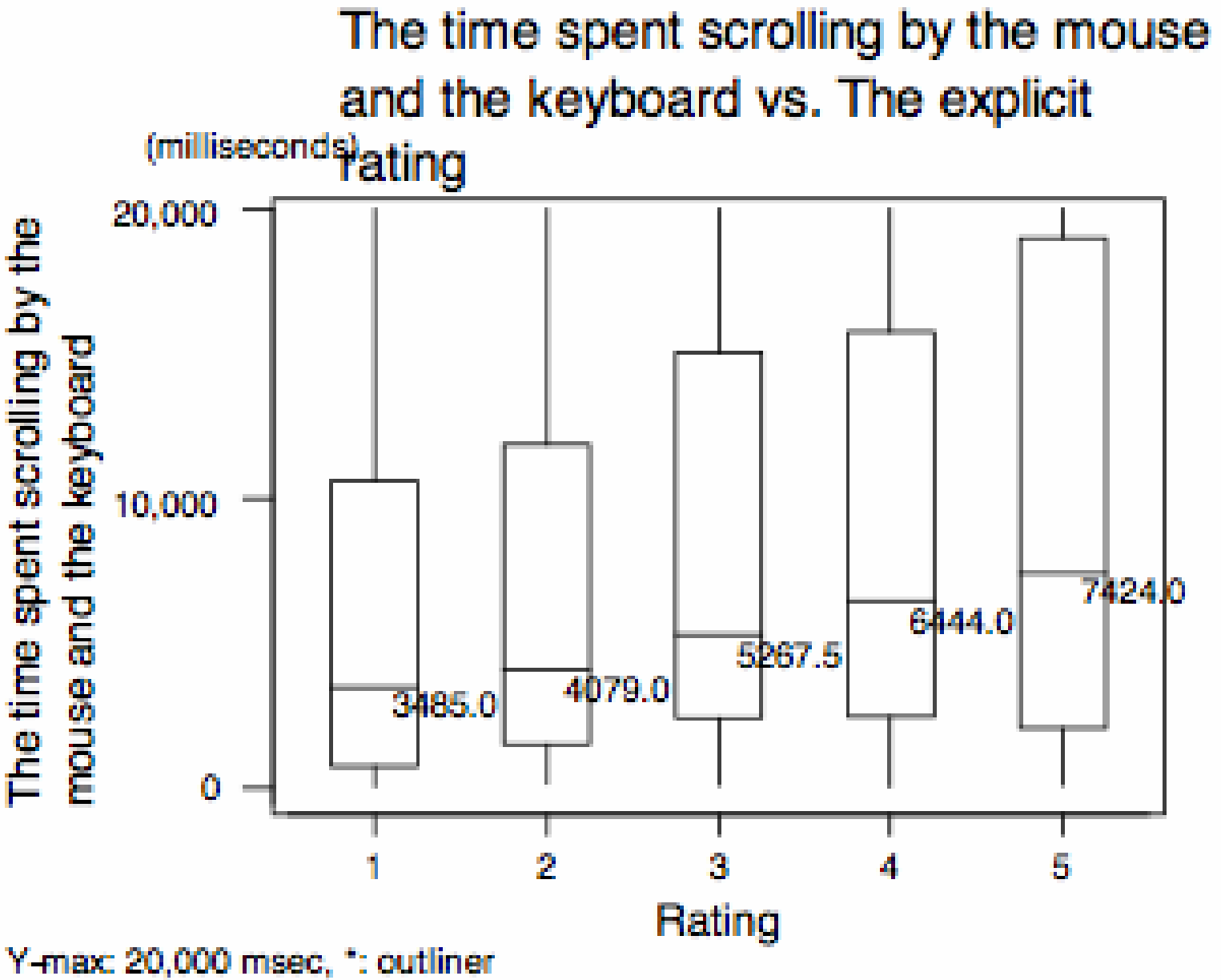
Rating

Based on the work of Claypool et al 2001

- The time a user spends on a website
- The time the cursor is in motion
- The number of mouse clicks
- The time a user scrolls

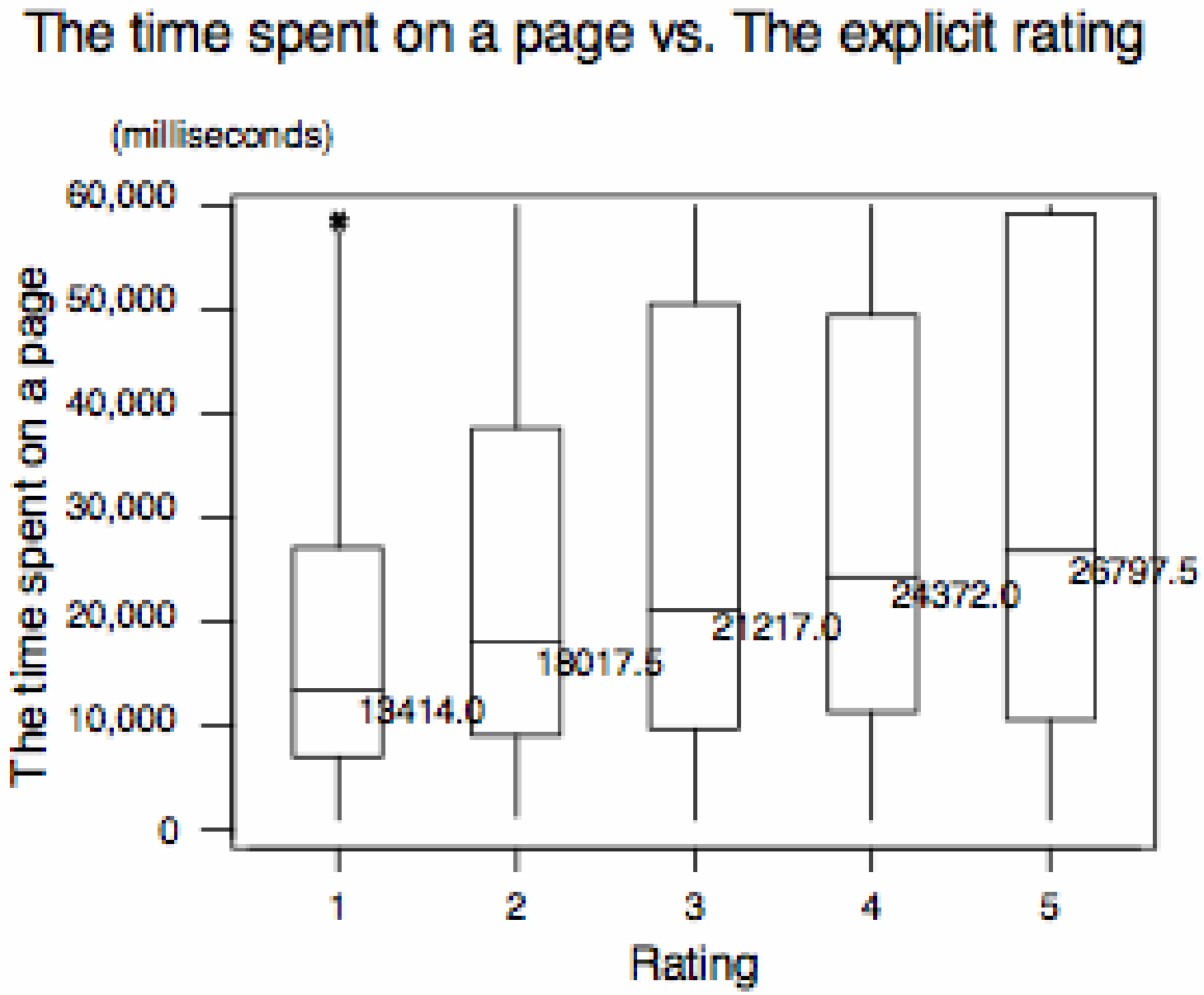
Rating

Claypool et al 2001



Rating

Claypool et al 2001



Y-max: 60,000 msec, *: outliner

Rating

$$p = \frac{TimeSpend}{TruncatedMeanOfTimeSpend}$$

Rating for Time a user spends				
1	2	3	4	5
$0\% \leq p < 76\%$	$76\% \leq p < 95\%$	$95\% \leq p < 110\%$	$110\% \leq p < 125\%$	-

Rating

$$p = \frac{TimeScrolled}{TruncatedMeanOfTimeScrolled}$$

Rating for Time a user scrolls				
1	2	3	4	5
$0\% \leq p < 69\%$	$69\% \leq p < 86\%$	$86\% \leq p < 108\%$	$108\% \leq p < 129\%$	$129\% \leq p$

Tagging

Find the keywords that describe the question.

Tagging

Find the keywords that describe the question.

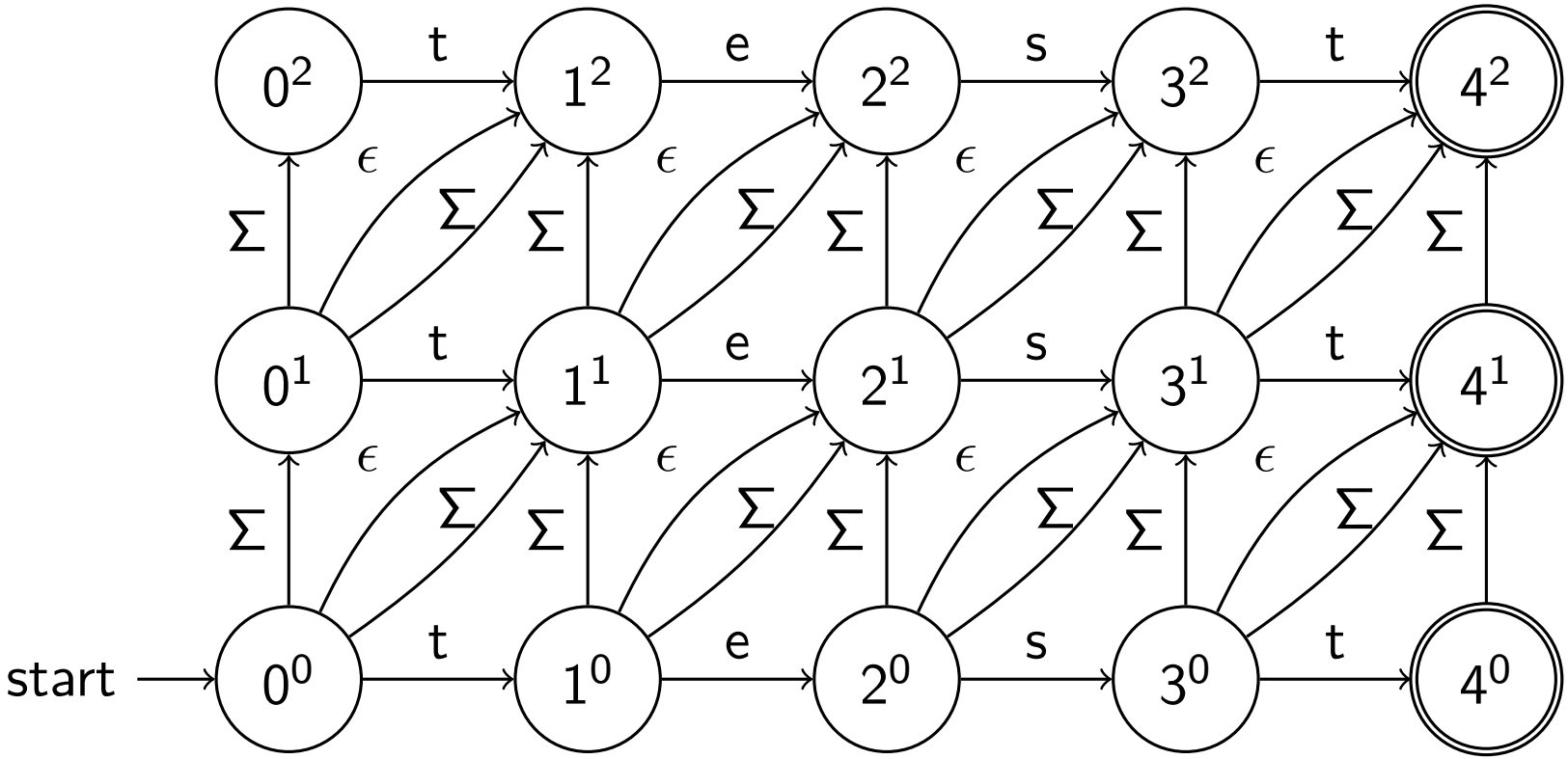
- STW Thesaurus for Economics

Tagging

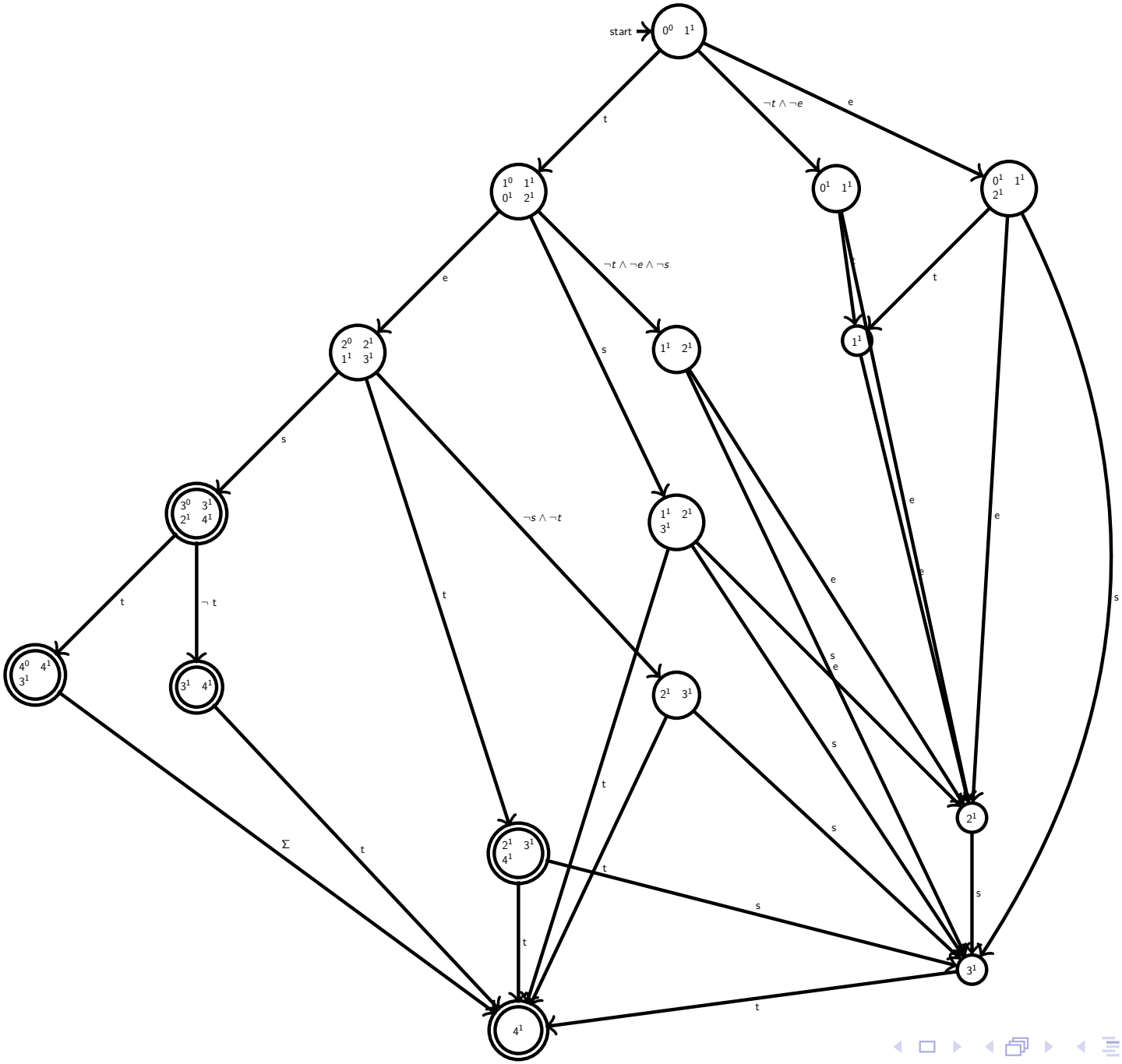
Find the keywords that describe the question.

- STW Thesaurus for Economics
- Computing Levenshtein distance: Calculate the distance between two words.
  Example:
  Libraries, Library
  Distance: 3

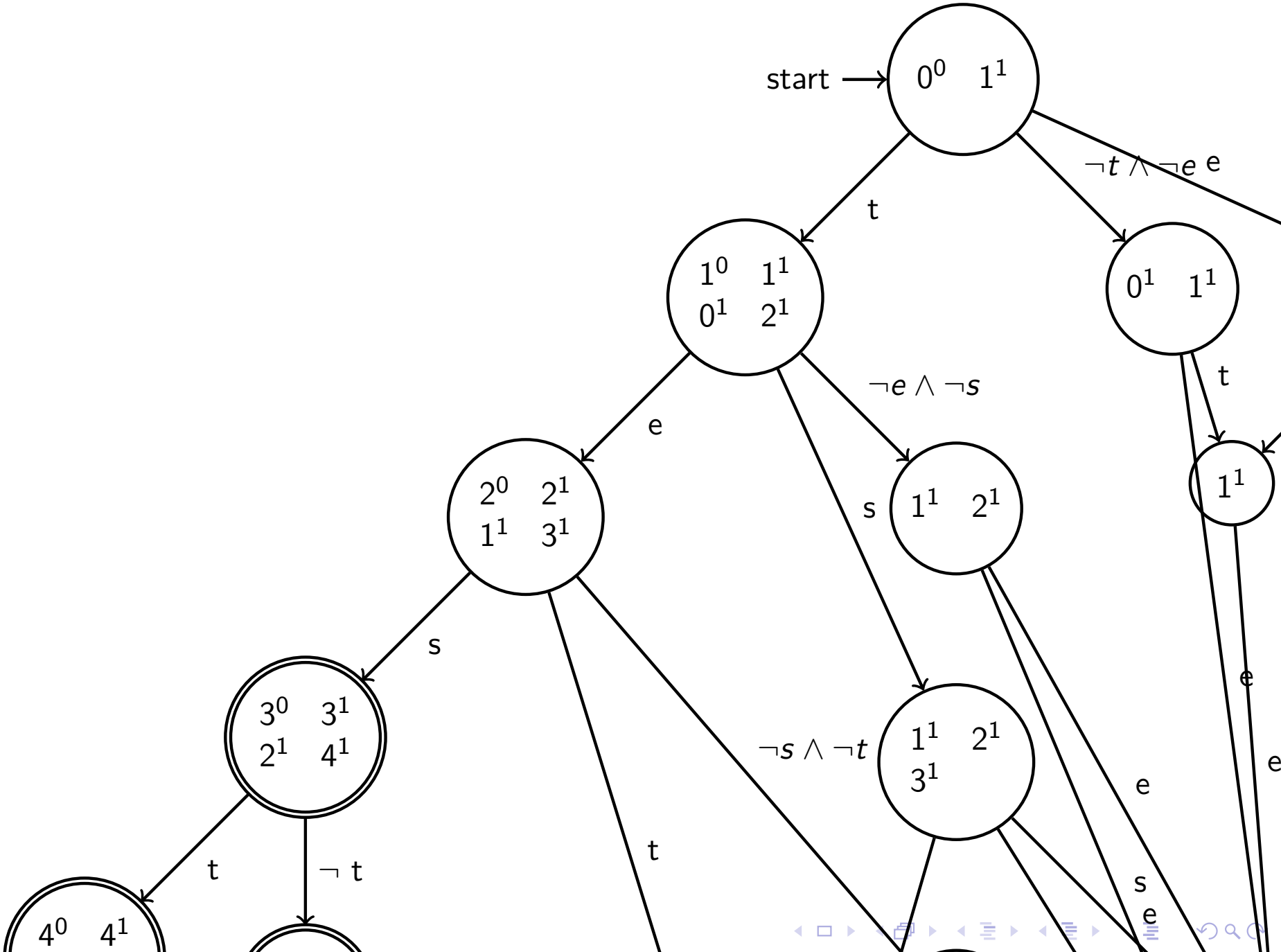
Levenshtein Automata



Levenshtein Automata



Levenshtein Automata



Maximum Distance for Words

Length	≤ 3	≤ 5	$6 \leq$
Distance	0	1	3

Maximum Distance for Words

Length	≤ 3	≤ 5	$6 \leq$
Distance	0	1	3

Problem Feminina with *e-plural: Hand Hände*

Maximum Distance for Words

Length	≤ 3	≤ 5	$6 \leq$
Distance	0	1	3

Problem Feminina with *e-plural: Hand Hände*
Maskulina und Neutra mit er-Plural 3 änderungen Wurm Würmer
Substantive mit er plural Fass Fässer
Based on *Die Pluralbildung im Deutschen: Eine Untersuchung an Hand der Optimalitätstheorie: "German Noun Plural reconsidered" von Dieter Wunderlich*

Tagging Evaluation

Tagging Algorithm Evaluation					
Words	STW	Non Matching	Matched Words	Levenshtein Time	Total Time
1	1	0	2	0.099	14.481
10	3	7	3	0.2843	20.348
100	30	70	54	0.4751	79.9616
1000	300	700	470	2.5606	777.4643

Item-Based Algorithm

	Item1	Item2	Item3	Item4	Item5
User1	5	3	4	4	?
User2	?	1	?	3	3
User3	4	?	4	3	5
User4	3	3	?	5	4
User5	1	5	5	2	1

Item-Based Algorithm

	Item1	Item2	Item3	Item4	Item5
User1	5	3	4	4	?
User2	?	1	?	3	3
User3	4	?	4	3	5
User4	3	3	?	5	4
User5	1	5	5	2	1

Cosinus Similarity

$a = [a_1, a_2, \ldots, a_n], b = [b_1, b_2, \ldots, b_n]$

$$\text{sim}(\vec{a}, \vec{b}) = \frac{\vec{a} \cdot \vec{b}}{||\vec{a}|| \cdot ||\vec{b}||} = \frac{\sum_{i=1}^{n} \vec{a_i} \vec{b_i}}{\sqrt{\sum_{i=1}^{n} \vec{a_i}^2} \sqrt{\sum_{i=1}^{n} \vec{b_i}^2}}$$

Take the differences of the average rating behaviour of the user into account.

Adjusted Cosinus Similarity

$$sim(a, b) = \frac{\sum_{u \in U} (r_{u,a} - \overline{r_u})(r_{u,b} - \overline{r_u})}{\sqrt{\sum_{u \in U} (r_{u,a} - \overline{r_u})^2} \sqrt{\sum_{u \in U} (r_{u,b} - \overline{r_u})^2}}$$

Predictions

Prediction

User u, Item p, Rating $r_{u,p}$

$$pred(u,p) = \frac{\sum_{i \in ratedItems(u)} sim(i,p) \cdot r_{u,i}}{\sum_{i \in ratedItems(u)} sim(i,p)}$$

Singular Value Decomposition

$$M = \begin{pmatrix} m_{11} & m_{12} & m_{13} \\ m_{21} & m_{22} & m_{23} \\ m_{31} & m_{32} & m_{33} \end{pmatrix}$$

Create a SVD with the matrix $M = U \cdot \Sigma \cdot V^t$

Singular Value Decomposition

$$M = \begin{pmatrix} m_{11} & m_{12} & m_{13} \\ m_{21} & m_{22} & m_{23} \\ m_{31} & m_{32} & m_{33} \end{pmatrix}$$

Create a SVD with the matrix $M = U \cdot \Sigma \cdot V^t$

$$\begin{pmatrix} u_{11} & u_{12} & u_{13} \\ u_{21} & u_{22} & u_{23} \\ u_{31} & u_{32} & u_{33} \end{pmatrix}$$

corresponds to the
column vectors of matrix
m

Singular Value Decomposition

$$M = \begin{pmatrix} m_{11} & m_{12} & m_{13} \\ m_{21} & m_{22} & m_{23} \\ m_{31} & m_{32} & m_{33} \end{pmatrix}$$

Create a SVD with the matrix $M = U \cdot \Sigma \cdot V^t$

$$\begin{pmatrix} u_{11} & u_{12} & u_{13} \\ u_{21} & u_{22} & u_{23} \\ u_{31} & u_{32} & u_{33} \end{pmatrix}$$

corresponds to the column vectors of matrix m

$$\begin{pmatrix} \sigma_{11} & 0 & 0 \\ 0 & \sigma_{22} & 0 \\ 0 & 0 & \sigma_{33} \end{pmatrix}$$

diagonal matrix with $\sigma_{ii} > 0$ and $\sigma_{ii} \geq \sigma_{i+1i+1}$

Singular Value Decomposition

$$M = \begin{pmatrix} m_{11} & m_{12} & m_{13} \\ m_{21} & m_{22} & m_{23} \\ m_{31} & m_{32} & m_{33} \end{pmatrix}$$

Create a SVD with the matrix $M = U \cdot \Sigma \cdot V^t$

$$\begin{pmatrix} u_{11} & u_{12} & u_{13} \\ u_{21} & u_{22} & u_{23} \\ u_{31} & u_{32} & u_{33} \end{pmatrix}$$

corresponds to the column vectors of matrix m

$$\begin{pmatrix} \sigma_{11} & 0 & 0 \\ 0 & \sigma_{22} & 0 \\ 0 & 0 & \sigma_{33} \end{pmatrix}$$

diagonal matrix with $\sigma_{ii} > 0$ and $\sigma_{ii} \geq \sigma_{i+1i+1}$

$$\begin{pmatrix} v_{11} & v_{12} & v_{13} \\ v_{21} & v_{22} & v_{23} \\ v_{31} & v_{32} & v_{33} \end{pmatrix}$$

corresponds to the row vectors of matrix m

Low Rank Approximation of M

$$\begin{pmatrix} m_{11} & m_{12} & m_{13} \\ m_{21} & m_{22} & m_{23} \\ m_{31} & m_{32} & m_{33} \end{pmatrix} = \begin{pmatrix} u_{11} & u_{12} & u_{13} \\ u_{21} & u_{22} & u_{23} \\ u_{31} & u_{32} & u_{33} \end{pmatrix} \cdot \begin{pmatrix} \sigma_{11} & 0 & 0 \\ 0 & \sigma_{22} & 0 \\ 0 & 0 & \sigma_{33} \end{pmatrix} \cdot \begin{pmatrix} v_{11} & v_{21} & v_{31} \\ v_{12} & v_{22} & v_{32} \\ v_{13} & v_{23} & v_{33} \end{pmatrix}$$

- Derive from Σ the matrix Σ_k (with k new rank of M) formed by replacing σ_{ii} with $i > k$ by zeros.
- Compute and output $M_k = U \cdot \Sigma_k \cdot V^T$ as the rank-k approximation to M.

Recommendation Thesis

	Item1	Item2	Item3	Item4	Item5
User1	5	3	4	4	?
User2	?	1	?	3	3
User3	4	?	4	3	5
User4	3	3	?	5	4
User5	1	5	5	2	1

Recommendation Evaluation

Recommendation Comparison 80.000 ratings 20.000 predictions			
Technology	MAE	Offline Computation Time	Prediction Time
Item Based	0.83187	3187.69	4208.49
SVD Concept	0.77937	32471.93	4315.31
SVD Three	0.79545	469.82	4395.75
SVD Average	0.78622	507.90	4226.31
Three	1.03333	0.00	0.49
Average	0.82714	112.13	249.18

Service Oriented Architecture

- RecommendationService
- ItemBasedService
- SVDBasedService
- TaggerService
- QuerySTWService
- RatingService
- WebService