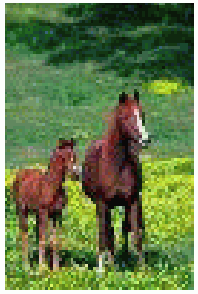
Evaluating Efficiency vs. Effectiveness for Vague Queries and Similarity Search in Digital Libraries

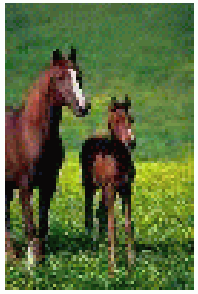
Norbert Fuhr

University of Dortmund, Germany

June 3, 2002

Motivating example


6282 0.00 5


6265 12.70 5

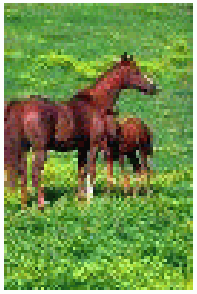

6278 13.86 5

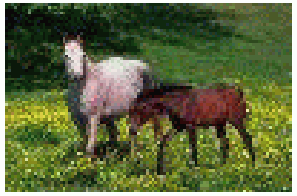

6288 14.18 5


6283 14.43 6


6245 15.97 4

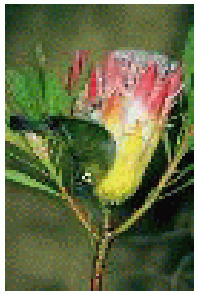

5737 16.30 5

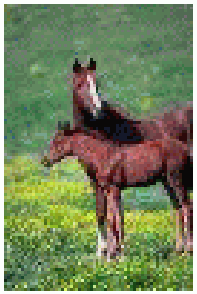

6261 16.43 4


6229 16.93 5

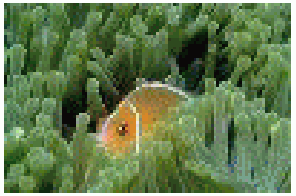

6296 17.09 3


6243 17.46 4


35280 17.55 6


6253 17.80 4


6239 18.03 4


35465 18.05 3

Effectiveness

$$\text{effectiveness} = \frac{\text{result quality}}{\text{user's effort}}$$

user's effort:

$\approx$ time for task completion

$=$ (response time + thinking time) * #iterations

Increasing effectiveness

$$\text{effectiveness} = \frac{\text{result quality}}{\text{user's effort}}$$

- increase result quality
  (may require more computations $\longrightarrow$ longer response times $\longrightarrow$ increased user effort)
- reduce user effort
  (reduce response time $\longrightarrow$ may decrease result quality)

$\longrightarrow$ tradeoff between result quality and response time

A new quality paradigm

example task: find 1 relevant document

I. 10 relevant documents in database,
system A succeeds 50% of the time

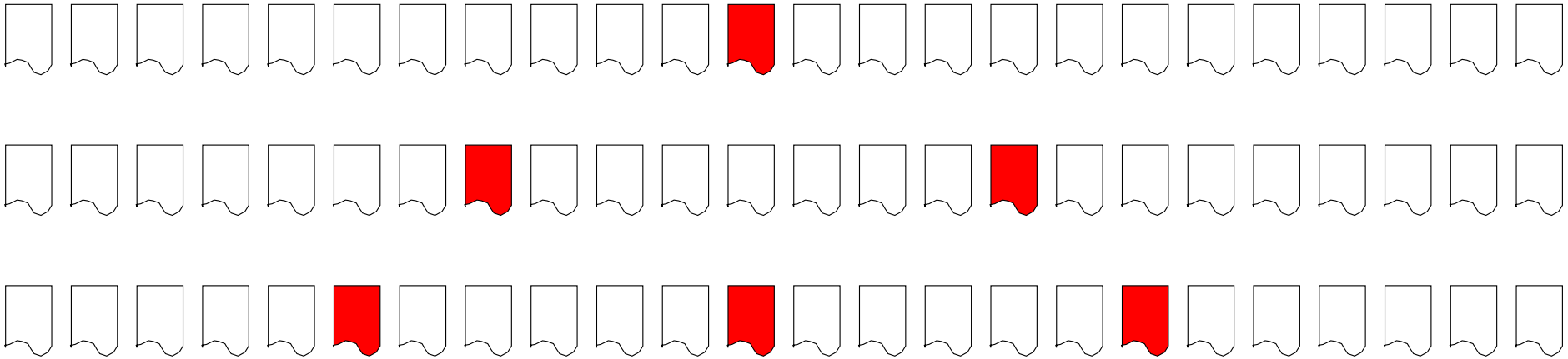
II. 100 relevant documents in database,
system B succeeds 50% of the time

Which system is better?

Ideal system

- perfect matching function
- linear scan through the database

task complexity = expected # processing steps for task completion



Task complexity

expected search length $\text{esl}(j, r, s) = \frac{j}{r+1} s$

r # relevant documents in collection

s # nonrelevant documents in collection

j # relevant documents requested

N collection size, $N \approx s$

task complexity $\text{tc}(j, r, N) = \frac{j}{r+1} N$

Example (cont'd)

I. 10 relevant documents in database,

system A succeeds 50% of the time

$$\text{tc}_I = \frac{1}{10+1}N$$

II. 100 relevant documents in database,

system B succeeds 50% of the time

$$\text{tc}_{II} = \frac{1}{100+1}N$$

$\longrightarrow$ system A does the harder task

Retrieval quality

define retrieval quality as relative quality:

$$\textbf{relative quality} = \frac{\text{complexity of task performed by real system}}{\text{complexity of task performed by ideal system}}$$

(standard recall and precision measures also compare performance with ideal system)

Find $\mathbf{n}$ documents

Task: User wants n documents

(e.g. Web retrieval: $n = 10$)

assume $r \geq n$ relevant documents in the database

- ideal system: retrieves only relevant docs

$$\mathsf{tc}_i = \frac{n}{r+1} N$$

- real system: finds only k relevant ones

$$\mathsf{tc}_r = \frac{k}{r+1} N$$

relative quality: $\dfrac{\mathsf{tc}_r}{\mathsf{tc}_i} = \dfrac{k}{n}$

$\rightarrow$ **Precision measure**

Find k relevant documents

Task: User wants k *relevant* documents

- ideal system: retrieves k documents, all relevant
- real system: retrieves n documents s.th. k relevant ones included (assumes all n documents are relevant)

$$\frac{\mathsf{tc}_r}{\mathsf{tc}_i} = \frac{\frac{k}{r \cdot n/k+1} N}{\frac{k}{r+1} N} = \frac{r+1}{r \cdot n/k+1} \approx \frac{k}{n}$$

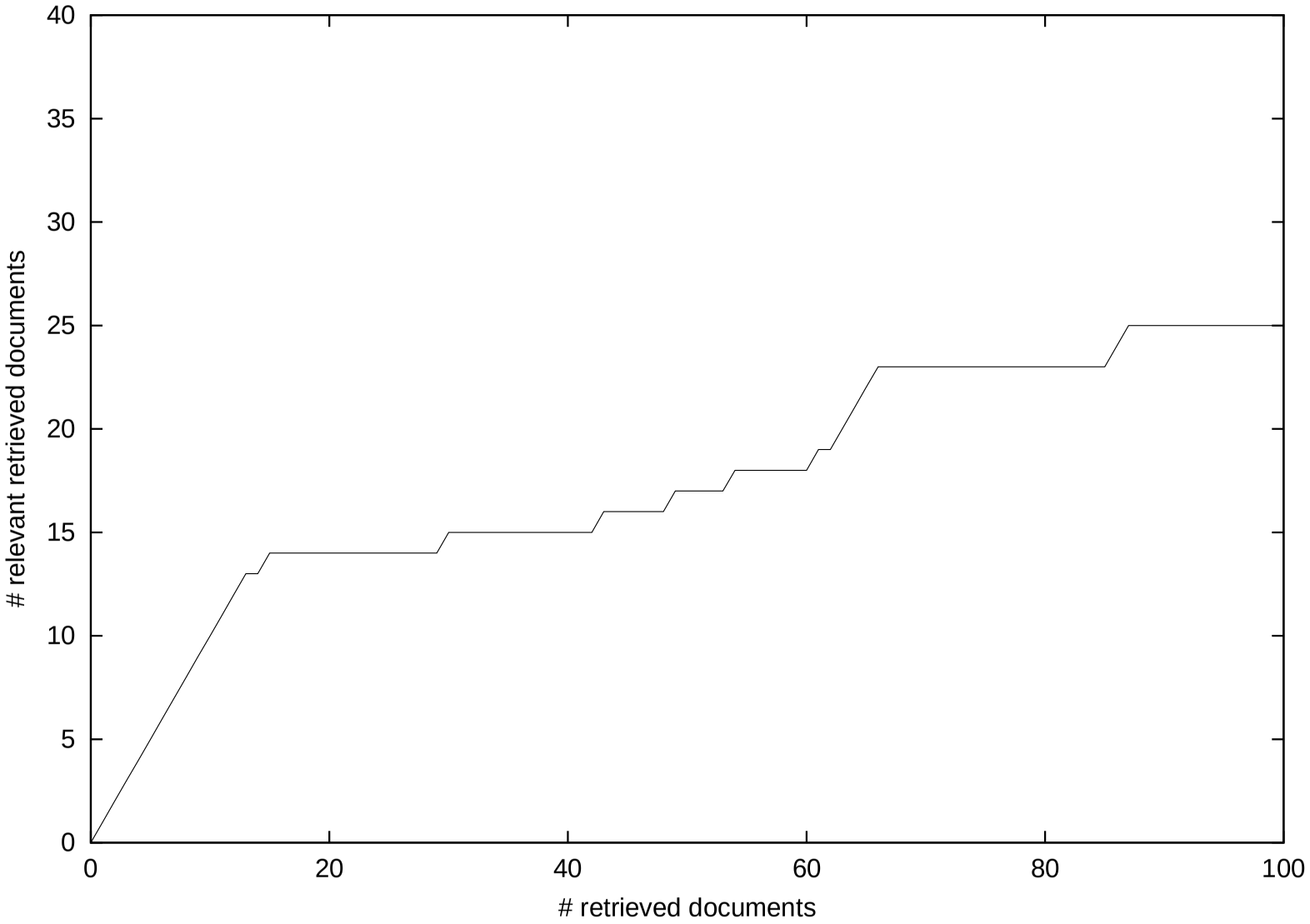
kNN search

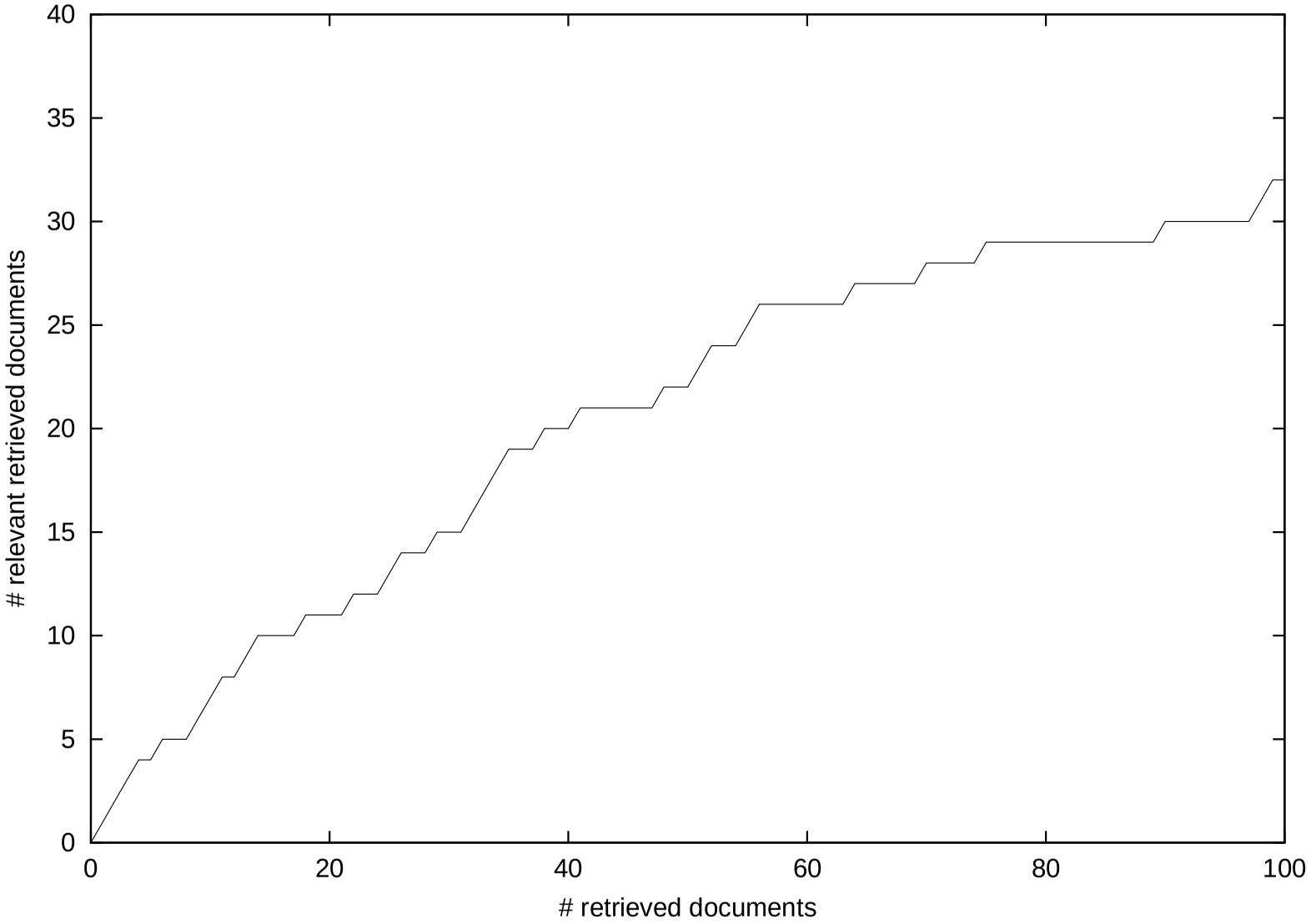
Task: User wants k closest matches

- ideal system: retrieves exactly k closest matches
- real system: retrieves k documents, kth document has real rank n
  $\rightarrow$ real system selects k documents from the top n ones

$$\frac{\mathsf{tc}_r}{\mathsf{tc}_i} = \frac{\frac{k}{n+1}N}{\frac{k}{k+1}N} = \frac{k+1}{n+1}$$

Similarity-based vs. relevance-based evaluation




40
35
30
25
20
15
10
5
0
relevant retrieved documents
0
20
40
60
80
100
retrieved documents

Summary and conclusions

- new paradigm for measuring retrieval quality
- $\longrightarrow$ new justification for recall and precision measure
- $\longrightarrow$ new measure for evaluating approximate kNN searches

Open issues

- User's perception of absolute quality?
- Application to complex structures in digital libraries?