information retrieval systems, beyond one-shot interaction and some research trends

jussi karlgren, gavagai and kth, stockholm

february 2019

last words from last time

1. reliability vs validity

last words from last time

1. reliability vs validity
2. relevance

last words from last time

1. reliability vs validity
2. relevance
3. precision and recall

last words from last time

1. reliability vs validity
2. relevance
3. precision and recall
4. some measures based on p & r

last words from last time

1. reliability vs validity
2. relevance
3. precision and recall
4. some measures based on p & r
5. averages are risky

what is quality in information access?

- usefulness (and effectiveness) for task

what is quality in information access?

- usefulness (and effectiveness) for task
- appealing presentation, response time, and intuitive interaction

what is quality in information access?

- usefulness (and effectiveness) for task
- appealing presentation, response time, and intuitive interaction
- authority, trustworthiness, sourceability, and reliability

what is quality in information access?

- usefulness (and effectiveness) for task
- appealing presentation, response time, and intuitive interaction
- authority, trustworthiness, sourceability, and reliability
- relevance and truthfulness

what is quality in information access?

- usefulness (and effectiveness) for task
- appealing presentation, response time, and intuitive interaction
- authority, trustworthiness, sourceability, and reliability
- relevance and truthfulness
- reusability and cost

what is quality in information access?

- usefulness (and effectiveness) for task
- appealing presentation, response time, and intuitive interaction
- authority, trustworthiness, sourceability, and reliability
- relevance and truthfulness
- reusability and cost
- timeliness in content

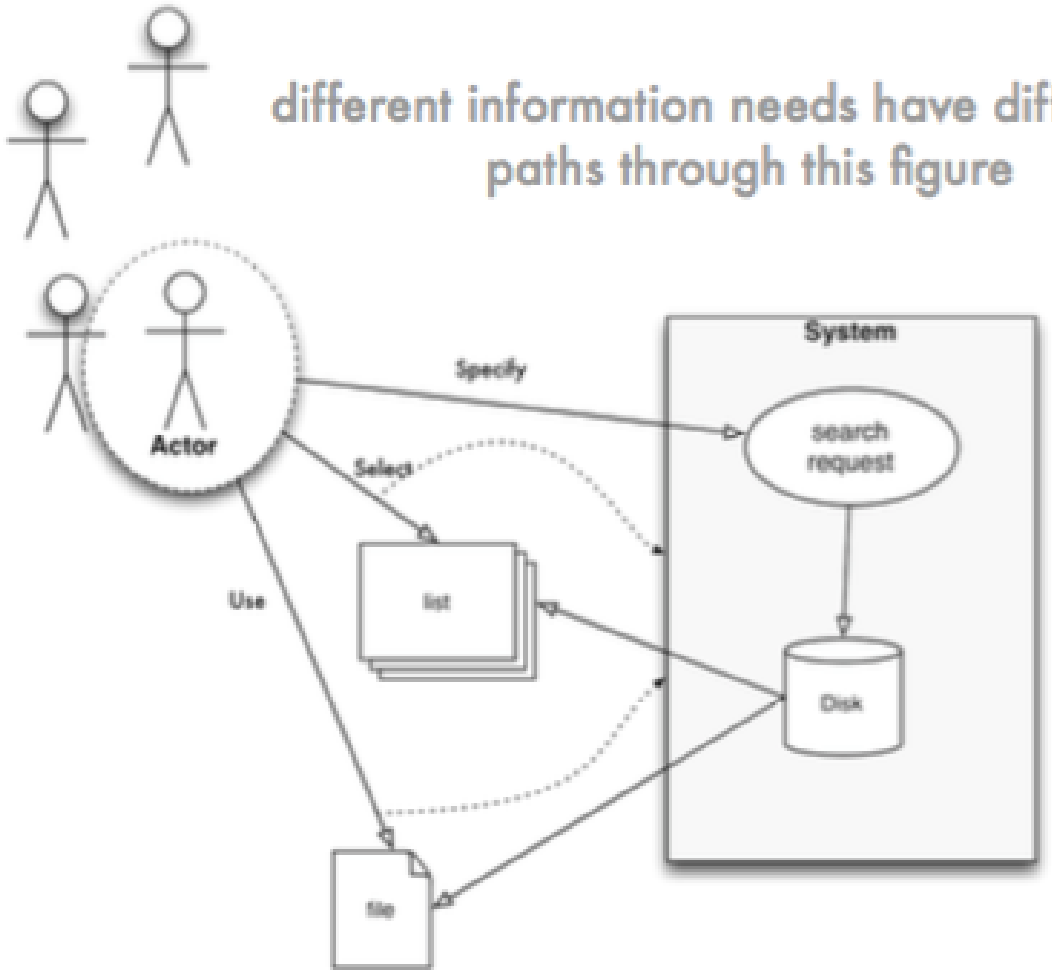
what is quality in information access?

- usefulness (and effectiveness) for task
- appealing presentation, response time, and intuitive interaction
- authority, trustworthiness, sourceability, and reliability
- relevance and truthfulness
- reusability and cost
- timeliness in content

how to measure nice things?

measuring depends on the envisioned situation of use

different information needs have different paths through this figure



improving systems

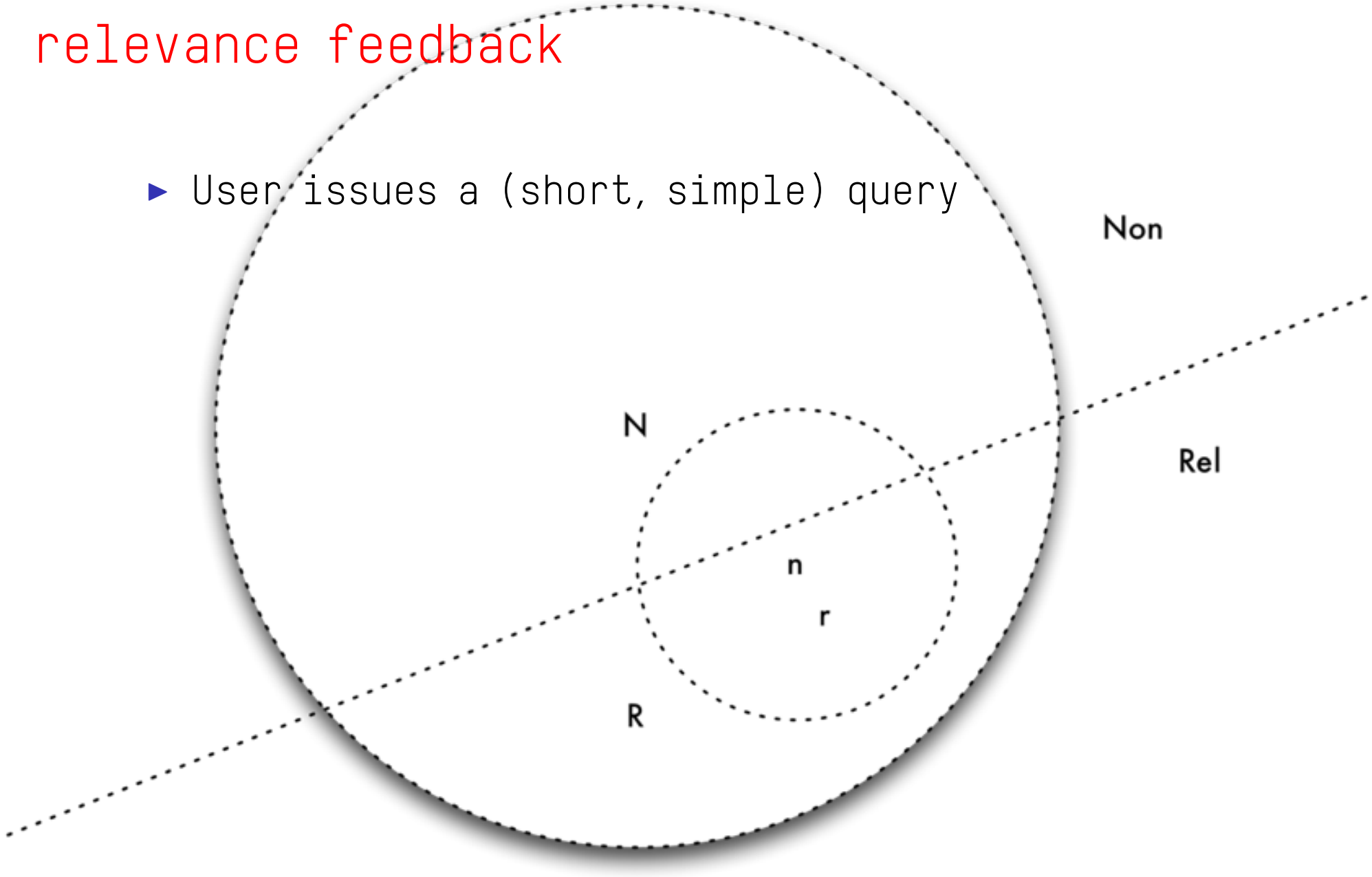
inspect what the user prefers

sessions!

use interaction
local: improve each session
global: improve across all sessions

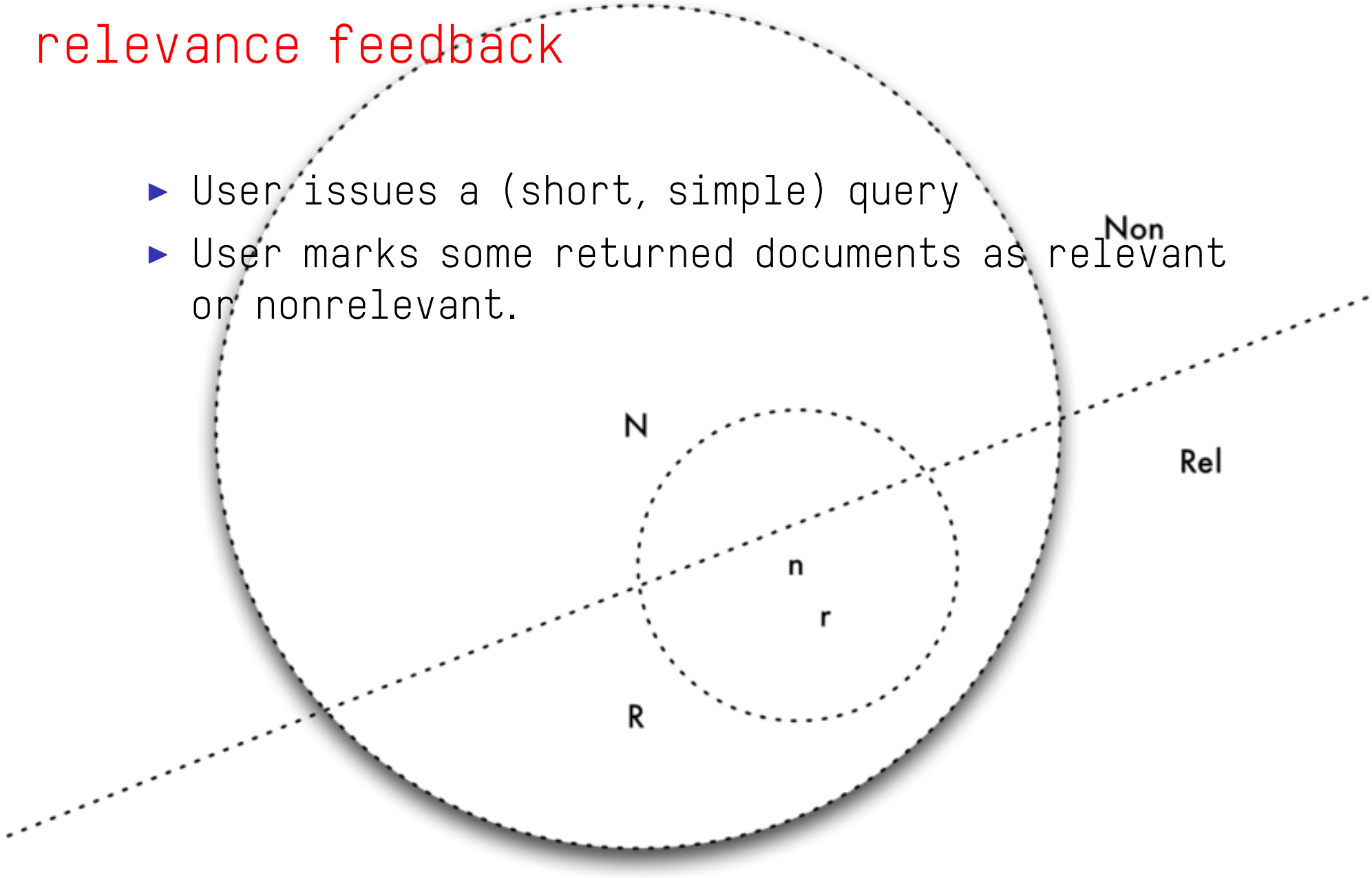
relevance feedback

- User issues a (short, simple) query



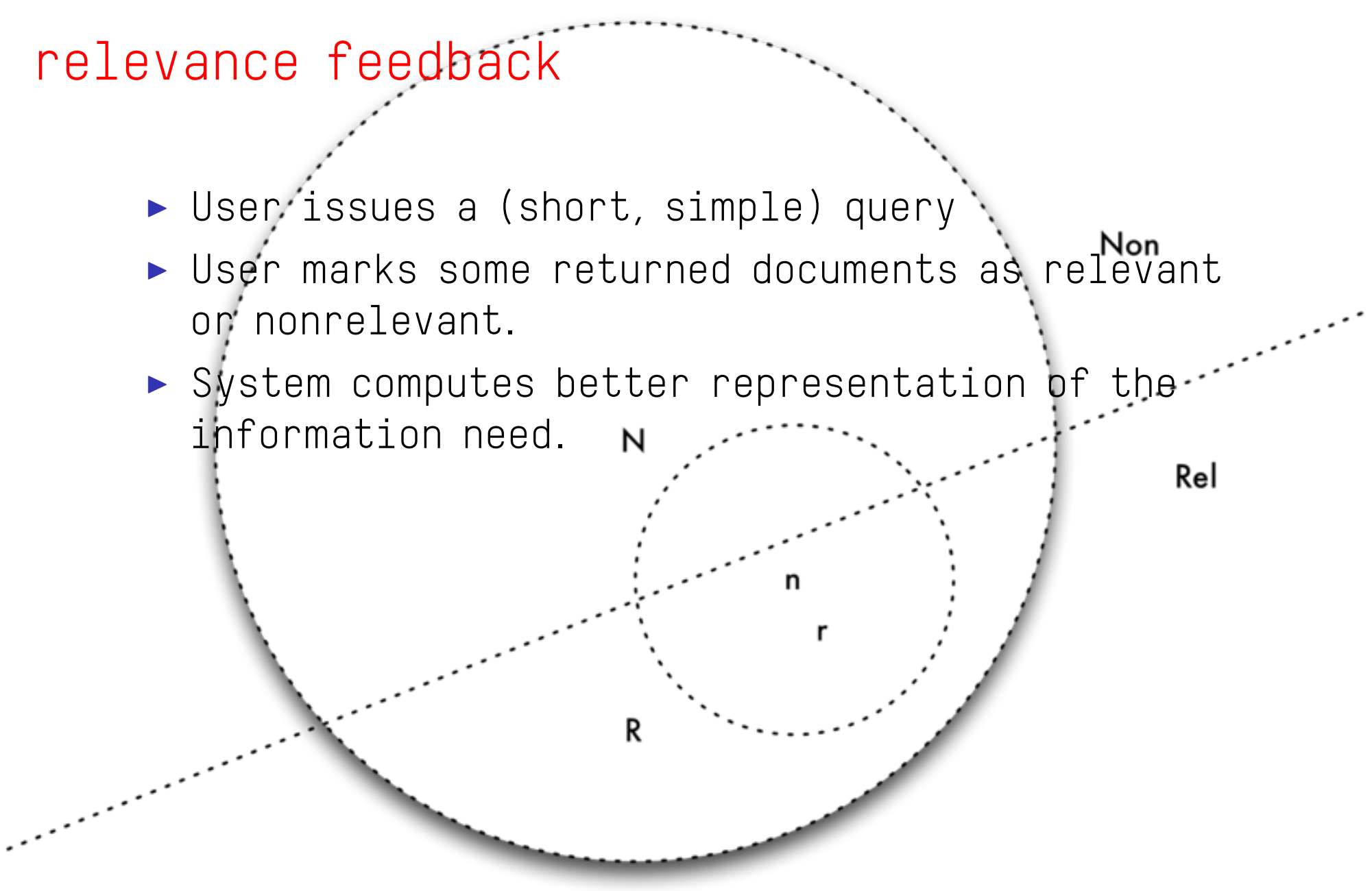
relevance feedback

- User issues a (short, simple) query
- User marks some returned documents as relevant or nonrelevant.



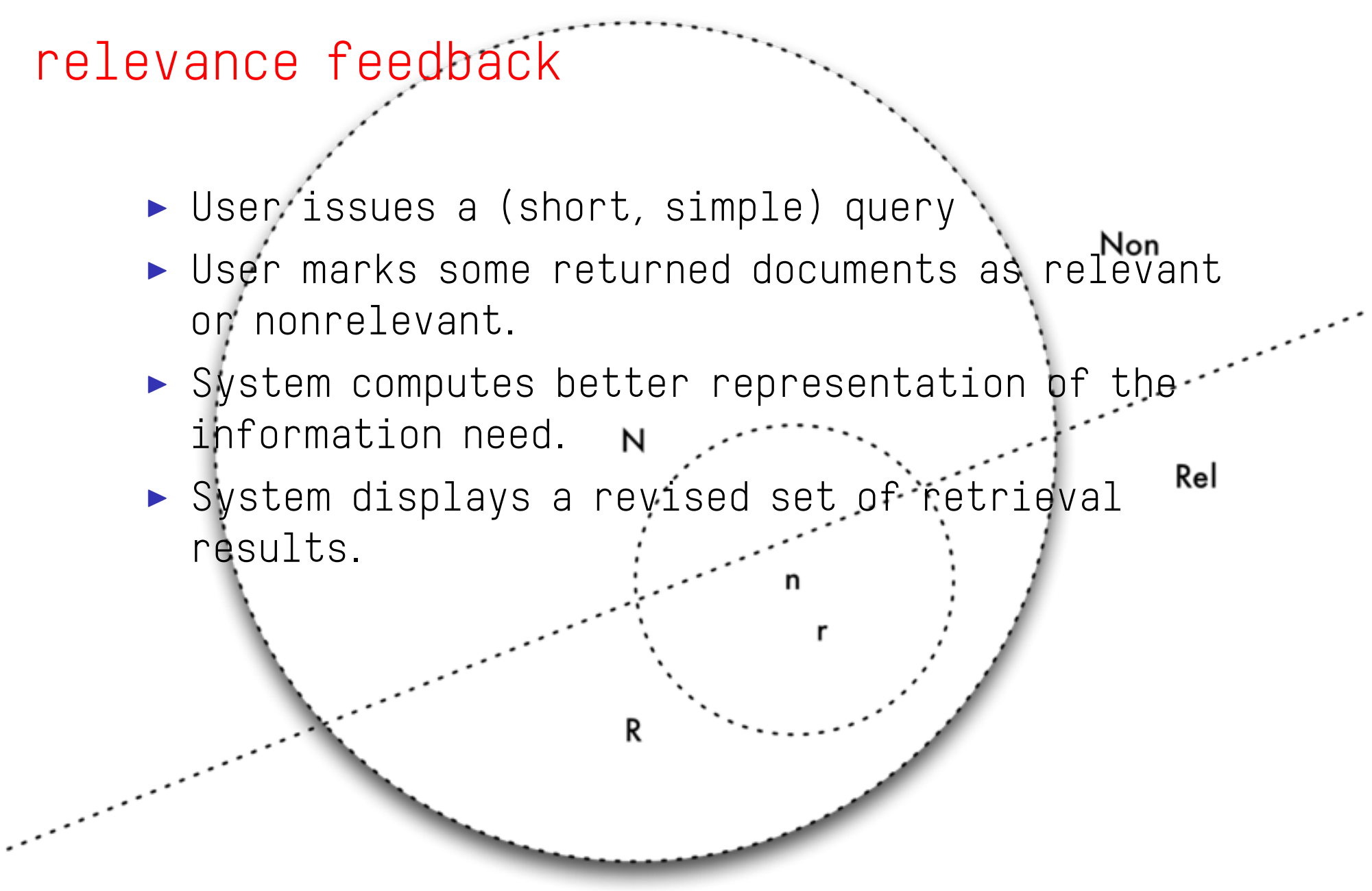
relevance feedback

- User issues a (short, simple) query
- User marks some returned documents as relevant or nonrelevant.
- System computes better representation of the information need.



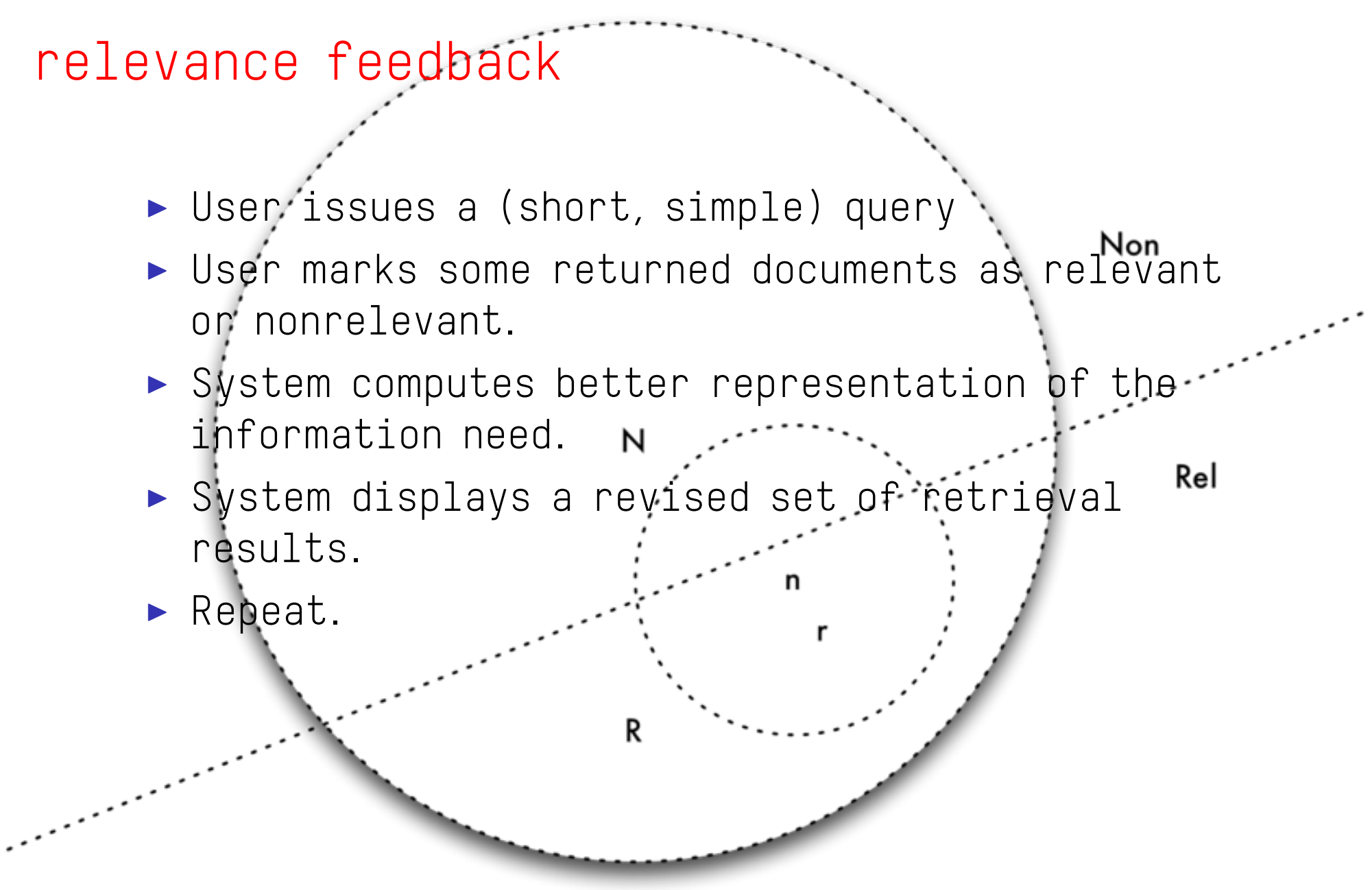
relevance feedback

- User issues a (short, simple) query
- User marks some returned documents as relevant or nonrelevant.
- System computes better representation of the information need.
- System displays a revised set of retrieval results.



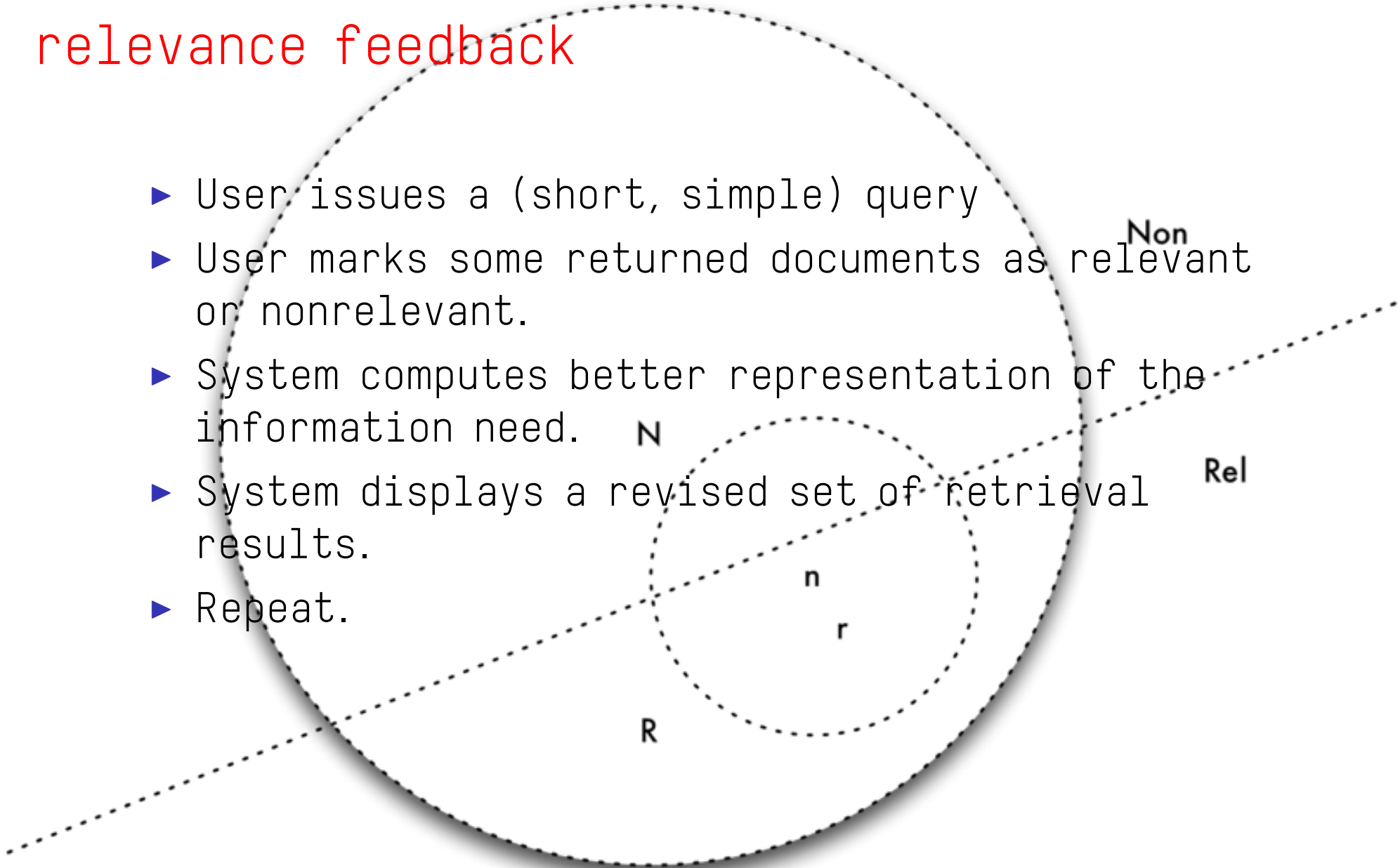
relevance feedback

- User issues a (short, simple) query
- User marks some returned documents as relevant or nonrelevant.
- System computes better representation of the information need.
- System displays a revised set of retrieval results.
- Repeat.



relevance feedback

- User issues a (short, simple) query
- User marks some returned documents as relevant or nonrelevant.
- System computes better representation of the information need.
- System displays a revised set of retrieval results.
- Repeat.



$$RW_{term} \simeq r/(n-r) \times ((N-n)-(R-r))/(R-r)$$

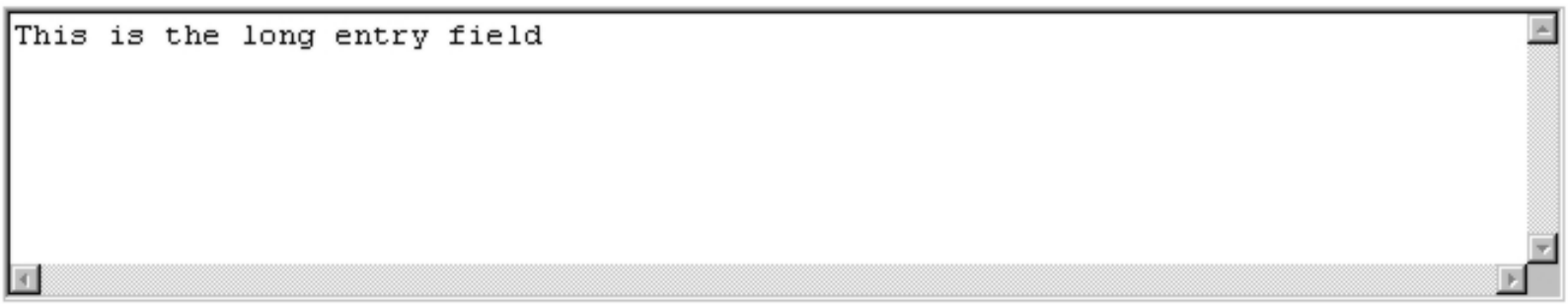
old slide

but we only get 1.87 wds / q

design matters for average query length

design matters for average query length

average 2.81 words



design matters for average query length

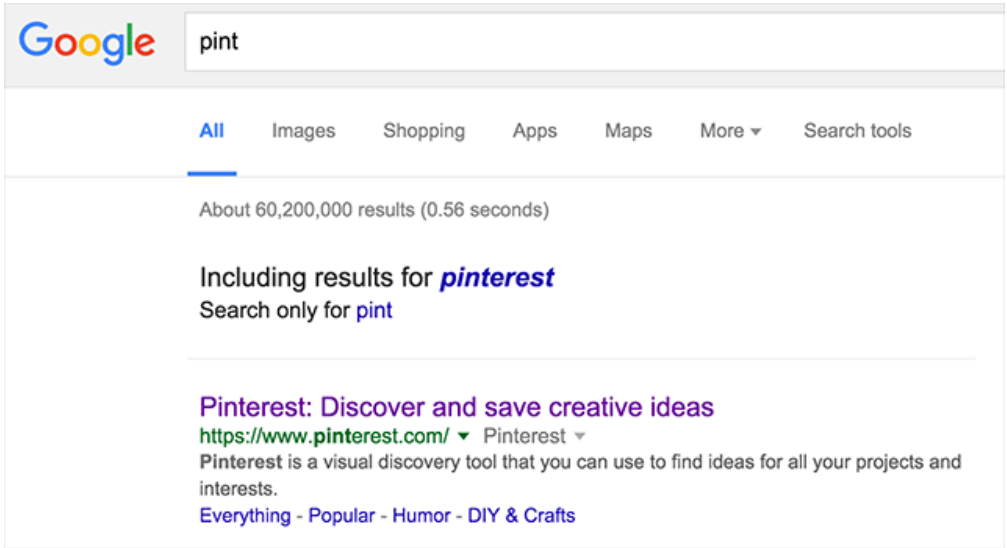
average 2.81 words

average 3.43 words

reweighting terms is useful for query expansion

friction

but users resist interaction



well... we have logs!

studiotraffic.com	Apr 1, 2006 14:39	1	http://www.studiotraffic.com
studiotraffic admin contact	Apr 2, 2006 9:19		
johndeere.com	Apr 2, 2006 11:12	1	http://www.deere.com
gurneys	Apr 3, 2006 15:31		
studio traffic	Apr 6, 2006 19:51	2	http://www.studiotraffic.org
studio traffic	Apr 6, 2006 19:51	3	http://www.scam.com
studio traffic	Apr 6, 2006 19:51	4	http://www.scam.com
what happened to studio traffic	Apr 6, 2006 19:57	1	http://www.talkgold.com
	...		
fort worth tx irving tx area hotels	Apr 26, 2006 11:40		
fort worth tx irving tx hotels	Apr 26, 2006 12:22		
buy cheap cds	Apr 28, 2006 22:36	1	http://www.cheap-cds.com
what is trimpatch 24-7	Apr 29, 2006 18:38	1	http://www.bodyandmindshop.co.uk
what is trimpatch 24-7	Apr 29, 2006 18:38	3	http://www.bodytherm.com
what is trimpatch 24-7	Apr 29, 2006 18:38	4	http://www.bodytherm.com
diet patch reviews	Apr 29, 2006 18:48	1	http://www.ultimatefatburner.com
diet patch reviews	Apr 29, 2006 18:48	4	http://www.reviewcentre.com
diet patch reviews	Apr 29, 2006 18:48	6	http://www.dietfraud.com
diet patch reviews	Apr 29, 2006 18:48	7	http://www.gamer-talk.net
diet patch reviews	Apr 29, 2006 18:48	9	http://www.nextag.com
treatment for double vision	Apr 30, 2006 9:40	4	http://www.medical-library.org
once a day diet pill	Apr 30, 2006 9:47	5	http://buyinnovations.com

more features of interest

log analysis
dwell time
click thru
returning visits
conversion rate
user actions
views

more features of interest

log analysis
dwell time
click thru
returning visits
conversion rate
user actions
views

what is a successful site? many visitors? long dwell time? most engagement?

apply this information to

apply this information to

- explicit reranking

apply this information to

- explicit reranking
- implicit reranking

apply this information to

- explicit reranking
- implicit reranking
- personalisation

old slide:

old slide:

old slide:

- user gets improved retrieval performance without interaction

old slide:

- user gets improved retrieval performance without interaction
- system performs normal retrieval

old slide:

- user gets improved retrieval performance without interaction
- system performs normal retrieval
- system assumes top k documents are relevant

old slide:

- user gets improved retrieval performance without interaction
- system performs normal retrieval
- system assumes top k documents are relevant
- system performs internal relevance feedback

old slide:

- user gets improved retrieval performance without interaction
- system performs normal retrieval
- system assumes top k documents are relevant
- system performs internal relevance feedback

new slide:

use machine learning: ``learning to rank''

gaming the search engines

https://news.artnet.com/market/
frieze-los-angeles-gretchen-andrew-1462594

let's try this

let's find a page and try to game a web search engine!

fun task-directed challenges

- diversity
- complex layouts
- dynamic planning processes in interaction step

fun technical challenges

- implicit feedback
- nonexistent feedback
- avoiding overfitting and working the decision boundary
- math is wrong for the very long tails

exasperating knowledge engineering challenges

- unclear target metric --- should be accessible, robust, differentiable (for most algorithms)
- legacy knowledge architectures
- distributed data challenges
- noisy feedback (did someone interrupt?)
- real-time vs training time issues
- pricing

``uncertainty is a first citizen''

important challenges with societal implications

- noisy data

important challenges with societal implications

- noisy data
- skewed data

important challenges with societal implications

- noisy data
- skewed data
- biased data

important challenges with societal implications

- noisy data
- skewed data
- biased data

FATE: fairness accountability transparency ethics (and confidentiality, and explainability ...)

``there are people on the other side of the interactive system you are building''

challenges going beyond adhoc retrieval

many tasks have retrieval at their core:

- chatbots, e.g.
- dialogue systems

challenges going beyond adhoc retrieval

many tasks have retrieval at their core:

- chatbots, e.g.
- dialogue systems

$\longrightarrow$

challenges going beyond adhoc retrieval

many tasks have retrieval at their core:

- chatbots, e.g.
- dialogue systems

$\longrightarrow$

- intent classification

challenges going beyond adhoc retrieval

many tasks have retrieval at their core:

- chatbots, e.g.
- dialogue systems

$\longrightarrow$

- intent classification
- task space

challenges going beyond adhoc retrieval

many tasks have retrieval at their core:

- chatbots, e.g.
- dialogue systems

$\longrightarrow$

- intent classification
- task space
- target metric

challenges going beyond adhoc retrieval

many tasks have retrieval at their core:

- chatbots, e.g.
- dialogue systems

$\longrightarrow$

- intent classification
- task space
- target metric
- multimodality

challenges going beyond adhoc retrieval

many tasks have retrieval at their core:

- chatbots, e.g.
- dialogue systems

$\longrightarrow$

- intent classification
- task space
- target metric
- multimodality
- entity linking

challenges going beyond adhoc retrieval

many tasks have retrieval at their core:

- chatbots, e.g.
- dialogue systems

$\longrightarrow$

- intent classification
- task space
- target metric
- multimodality
- entity linking
- urgency metrics (and other subjective language)

challenges going beyond adhoc retrieval

many tasks have retrieval at their core:

- chatbots, e.g.
- dialogue systems

$\longrightarrow$

- intent classification
- task space
- target metric
- multimodality
- entity linking
- urgency metrics (and other subjective language)
- developing without billion scale data

challenges going beyond adhoc retrieval

many tasks have retrieval at their core:

- chatbots, e.g.
- dialogue systems

$\longrightarrow$

- intent classification
- task space
- target metric
- multimodality
- entity linking
- urgency metrics (and other subjective language)
- developing without billion scale data
- moving from dev to billion scale data

best practice

satisficing vs optimisation

take home message

- information retrieval is a component

take home message

- information retrieval is a component
- sessions are more informative than one-shot

take home message

- information retrieval is a component
- sessions are more informative than one-shot
- relevance feedback is potentially useful

take home message

- information retrieval is a component
- sessions are more informative than one-shot
- relevance feedback is potentially useful
- log analysis can yield relevance feedback

take home message

- information retrieval is a component
- sessions are more informative than one-shot
- relevance feedback is potentially useful
- log analysis can yield relevance feedback
- analysis of current usage is inherently conservative

take home message

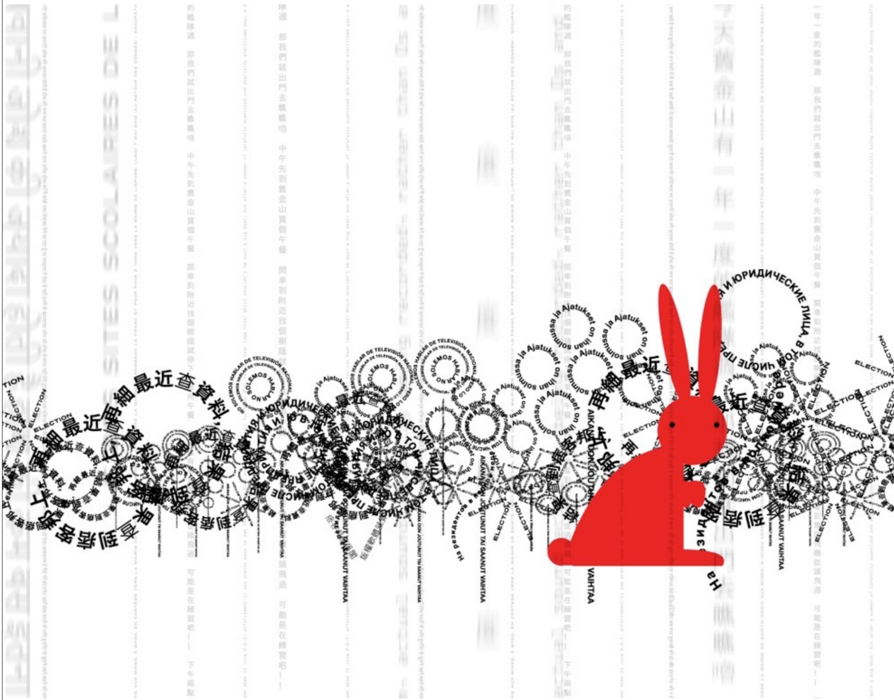
- information retrieval is a component
- sessions are more informative than one-shot
- relevance feedback is potentially useful
- log analysis can yield relevance feedback
- analysis of current usage is inherently conservative
- long tails are long and the future is longer than the past

take home message

- information retrieval is a component
- sessions are more informative than one-shot
- relevance feedback is potentially useful
- log analysis can yield relevance feedback
- analysis of current usage is inherently conservative
- long tails are long and the future is longer than the past
- new tasks require more sophisticated models than topic

take home message

- information retrieval is a component
- sessions are more informative than one-shot
- relevance feedback is potentially useful
- log analysis can yield relevance feedback
- analysis of current usage is inherently conservative
- long tails are long and the future is longer than the past
- new tasks require more sophisticated models than topic
- top line is not always best practice


И ЮРИДИЧЕСКИЕ ЛИЦА, В ТОМ ЧИСЛЕ
ELECTION
VAIHTAA
再細最近查資料