Improving Search Through *Efficient* A/B Testing:

A Case Study

Nokia Maps "Place Discovery" Team, Berlin:

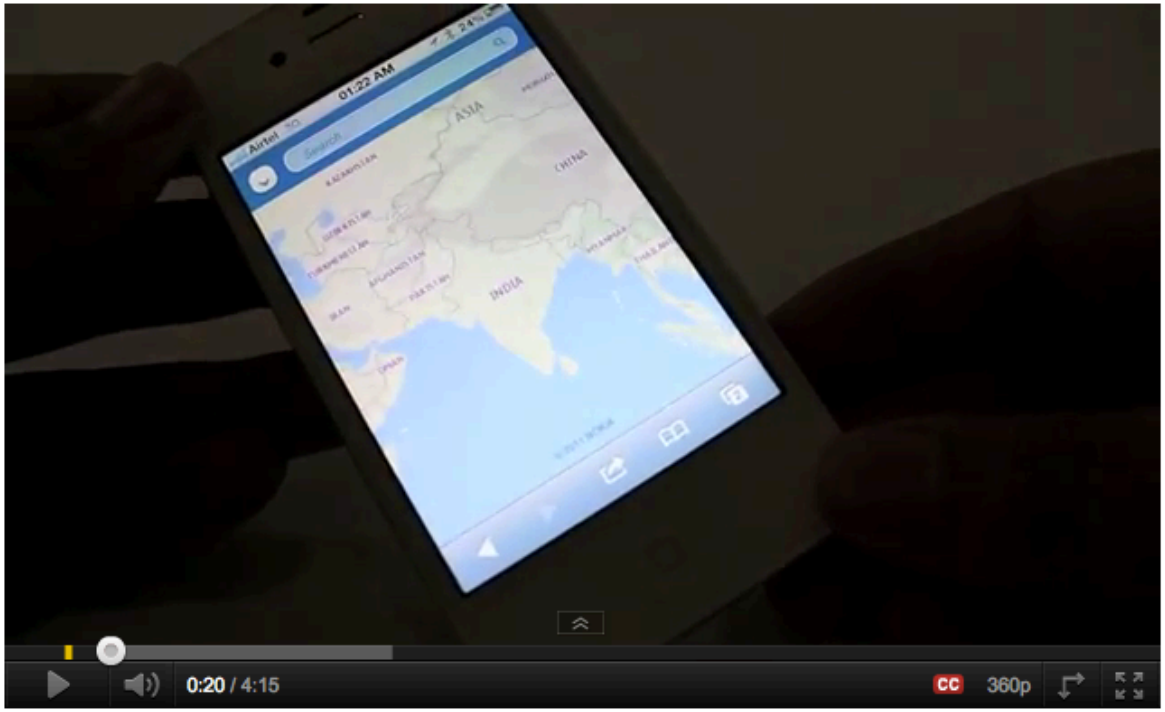
Hannes Kruppa, Steffen Bickel, Mark Waldaukat, Felix Weigel, Ross Turner, Peter Siemen

NOKIA

Nokia Maps for Everyone!

Nokia Maps on iPhone, iPad and Android

ZOMGitsCj.com 381 videos Subscribe



Like | Add to | Share

843 views

Uploaded by boybandreject00 on Jul 11, 2011

From http://zomgitscj.com/2011/07/12/nokia-maps-comes-to-iphone-ipad-and-android-b...

Nokia's Ovi Maps seems like it might make an appearance on iPhone, iPad

Show more

2 likes, 0 dislikes

As Seen On: Mobile Knots

Uploader Comments (boybandreject00)

Nokia 701 Hands-on Preview
by boybandreject00
9,953 views

Nokia N8 HD Video Sample - Driving Through the ...
by boybandreject00
1,669 views

Nokia 5233 Review : Camera UI
by boybandreject00
118,481 views

Nokia 600 Camera Video Sample
by boybandreject00

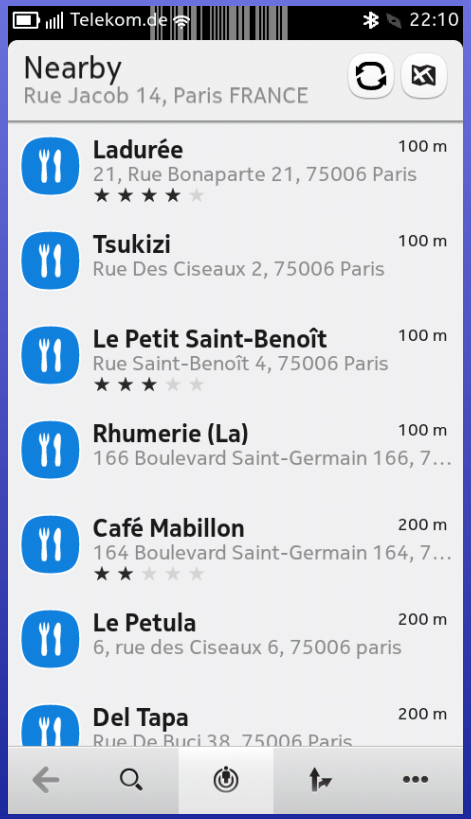
NOKIA

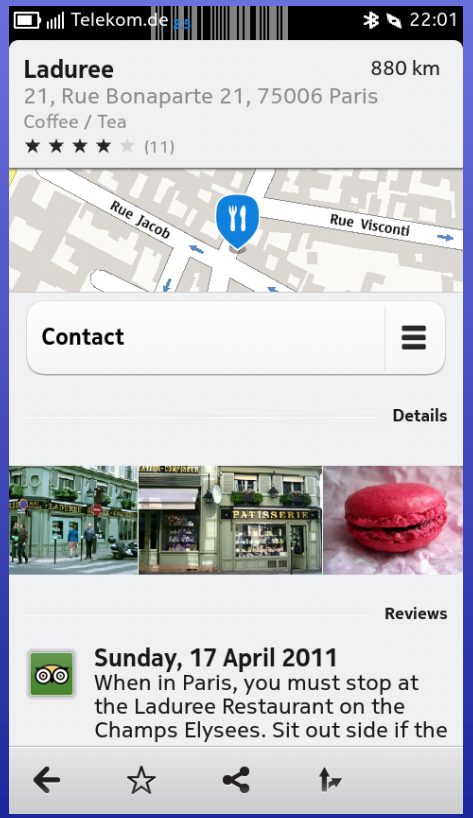
Nokia Maps Team, Berlin

NOKIA

Nokia Maps: *Nearby Places*

"Discover Places You Will Love, Anywhere"

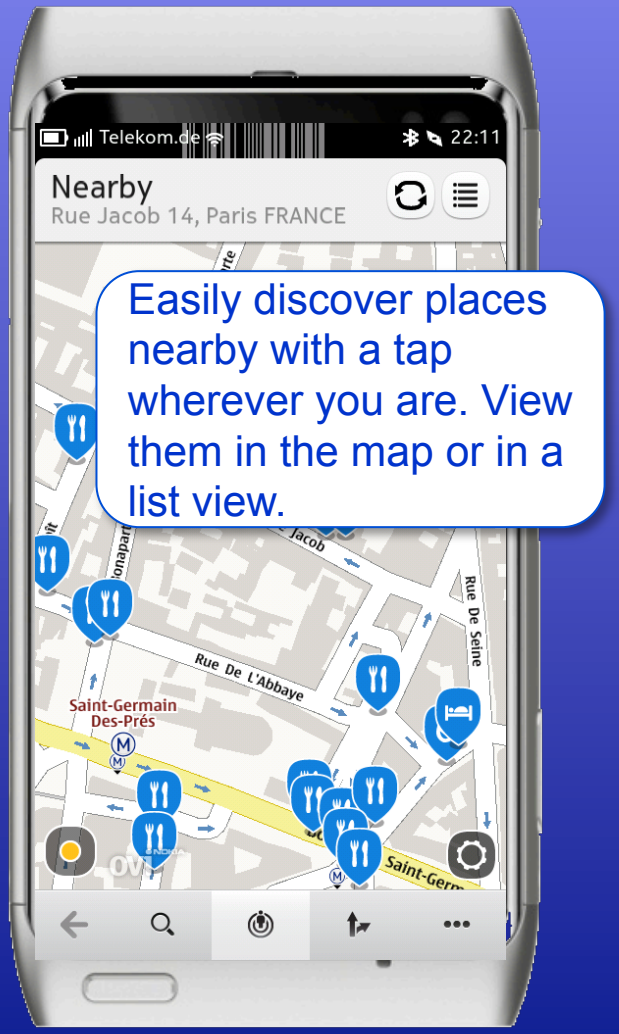


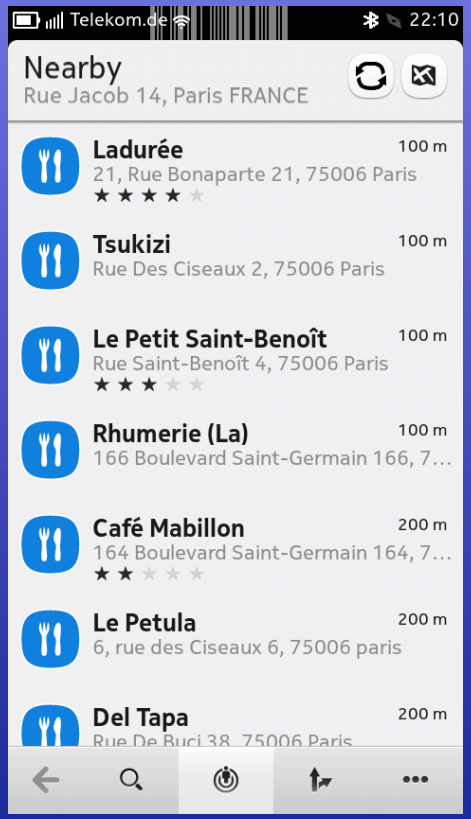


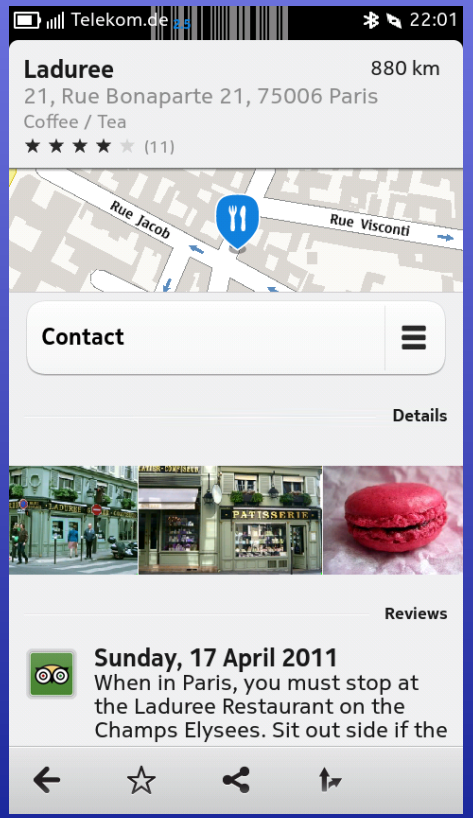
NOKIA

Nokia Maps: *Nearby Places*

"Discover Places You Will Love, Anywhere"



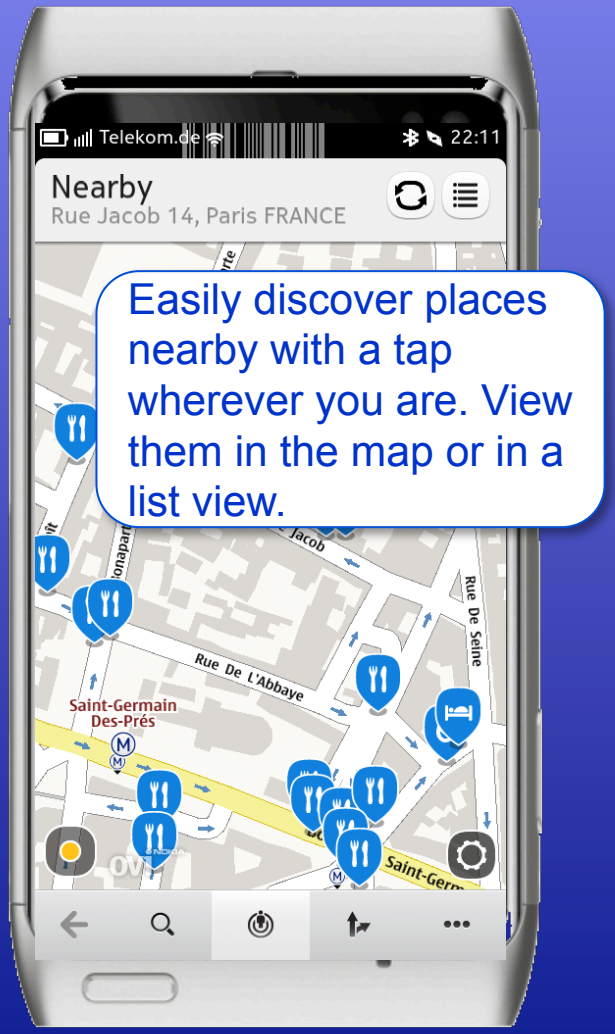


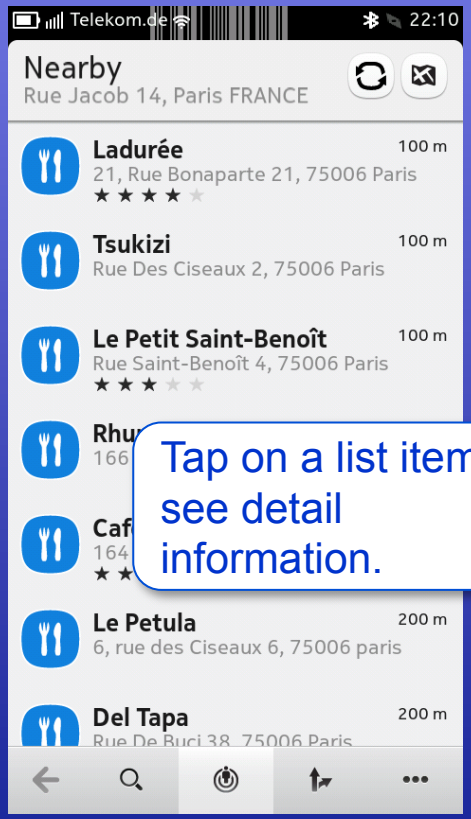


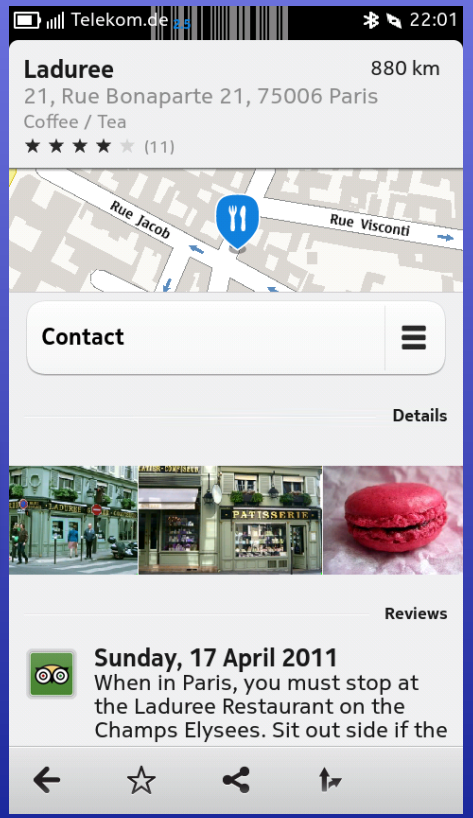
NOKIA

Nokia Maps: *Nearby Places*

"Discover Places You Will Love, Anywhere"



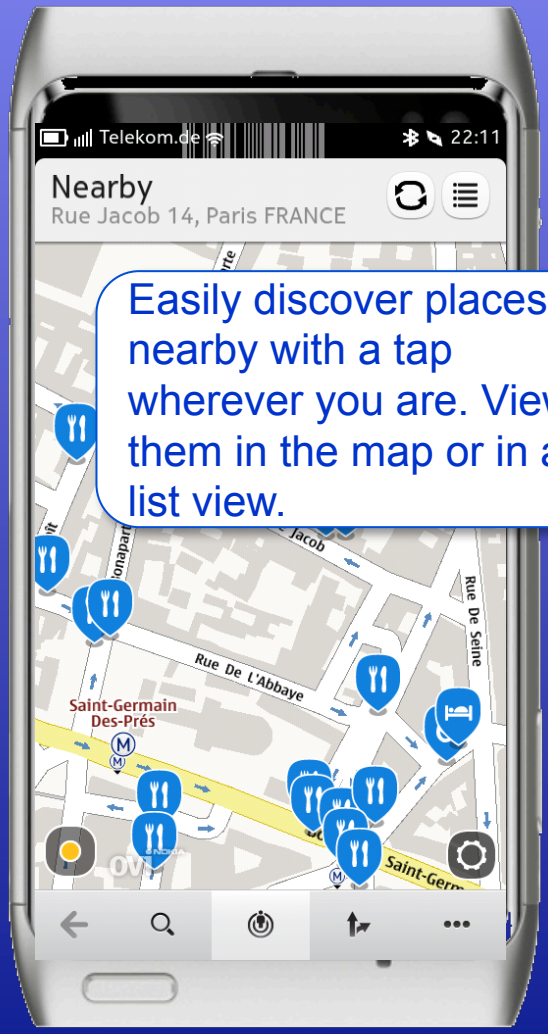




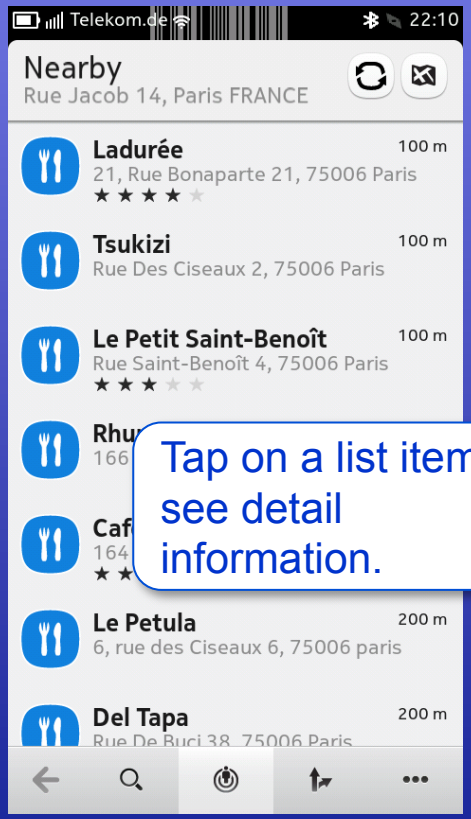
NOKIA

Nokia Maps: *Nearby Places*

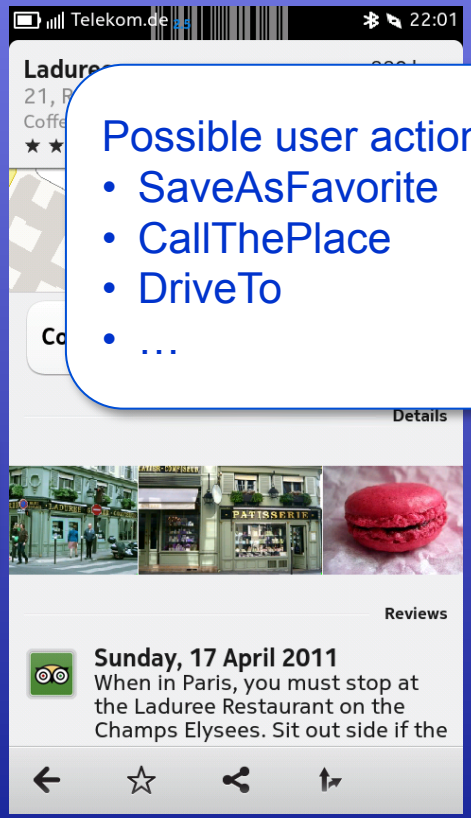
"Discover Places You Will Love, Anywhere"



Easily discover places nearby with a tap wherever you are. View them in the map or in a list view.



Tap on a list item to see detail information.



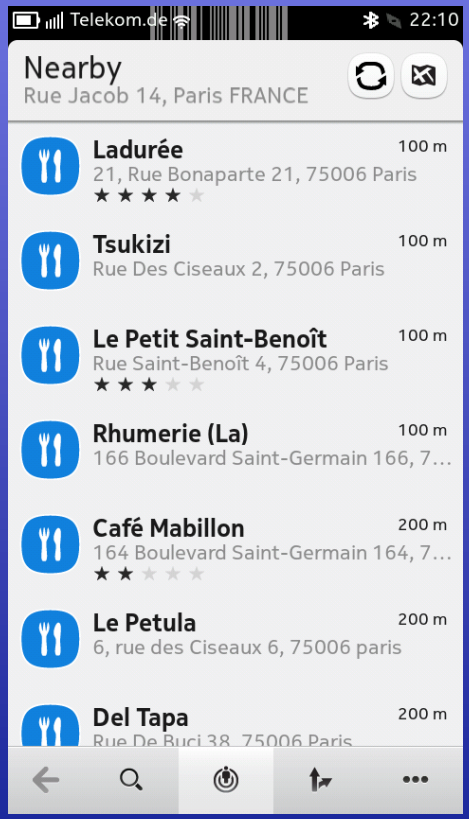
Possible user actions:

- SaveAsFavorite
- CallThePlace
- DriveTo
- …

NOKIA

Problem: Which Places to Show?

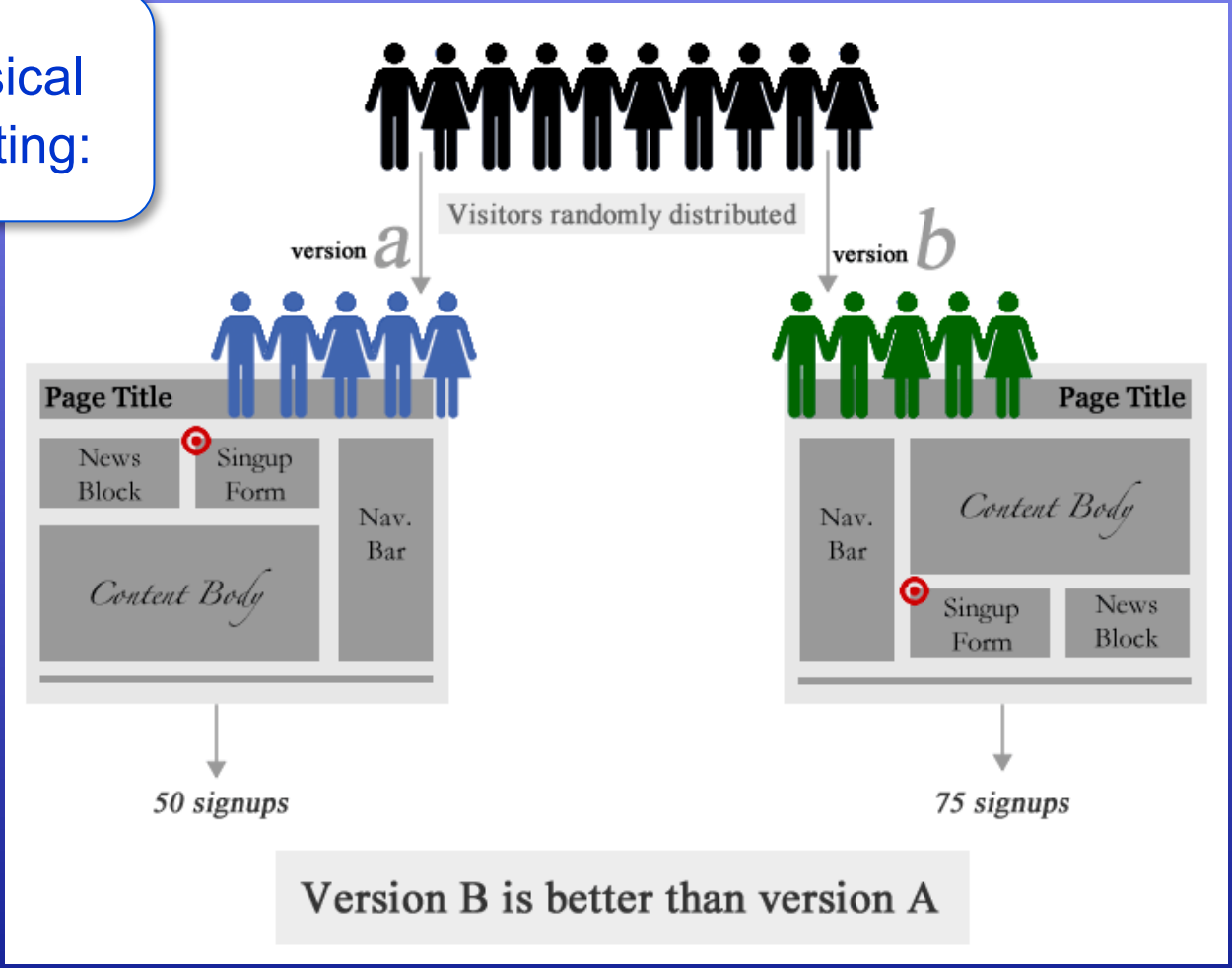
- Restaurants? Hotels? Shopping? ...
- rank by Ratings?
- Distance?
- Usage?
- Trending?
- ....



NOKIA

Approach: A/B-Test Different Versions!

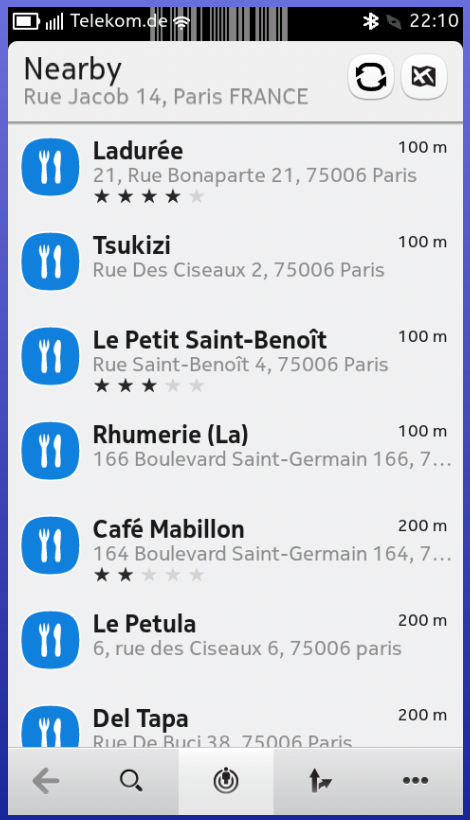
Here is classical Web A/B testing:



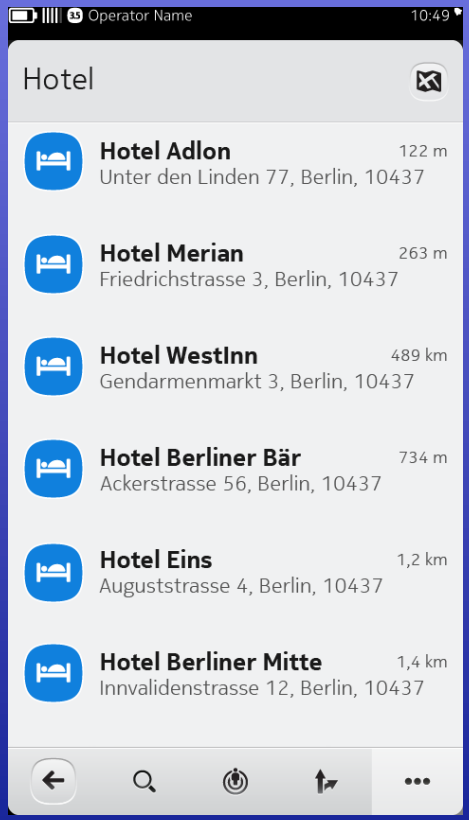
NOKIA

A/B-Test for *Nearby Places*

Version A:
Best of Eat'n'Drink



Version B:
Best of Hotels



Versions Compete for User engagement:

= *Number of Actions performed on places.*

NOKIA

There Is A Better Approach For Ranked Lists

[Joachims et al 2008]:

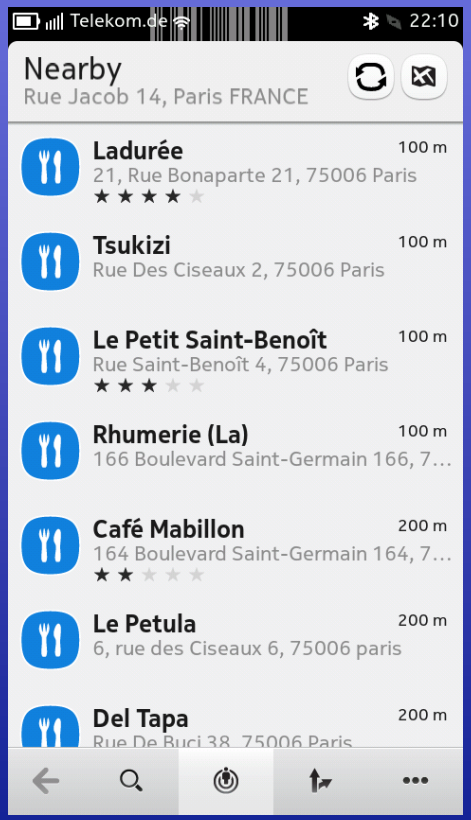
"How Does Clickthrough Data Reflect Retrieval Quality?"

- Classical A/B testing converges slowly for ranked lists
- Classical A/B testing often doesn't reflect actual relevance

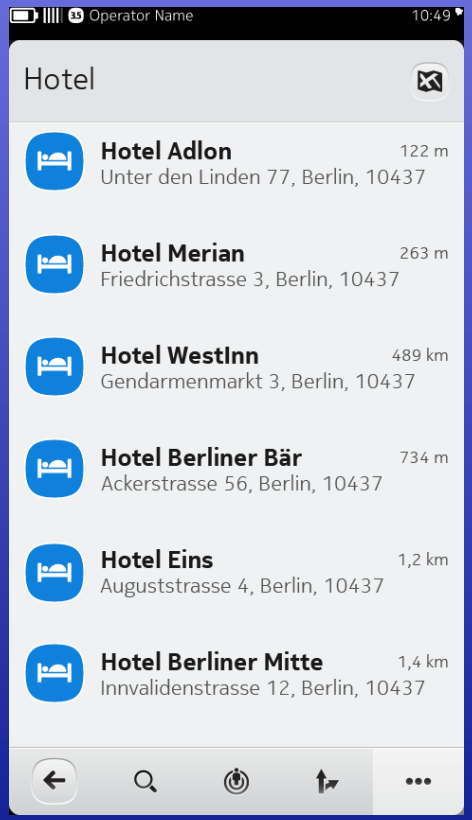
- A/B Tests for Ranked Result Lists: Rank- Interleaving
- Use Rank-Interleaving for faster statistical significance

Efficient A/B Testing: *Rank Interleaving*

Version A:
Best of Eat'n'Drink



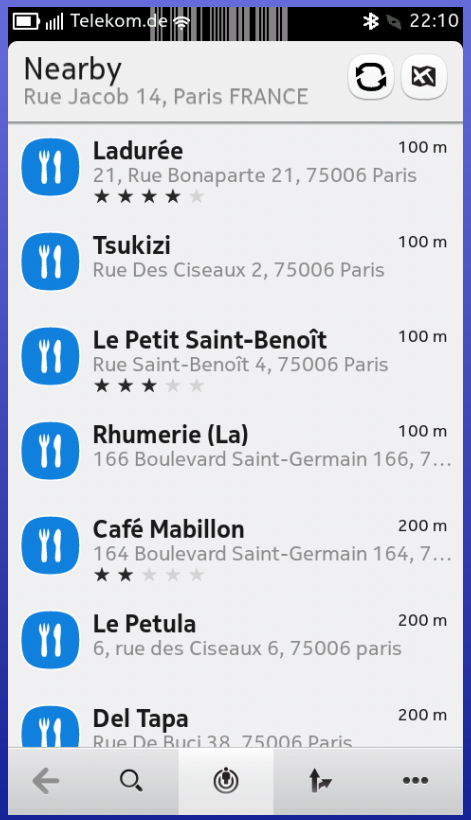
Version B:
Best of Hotels



NOKIA

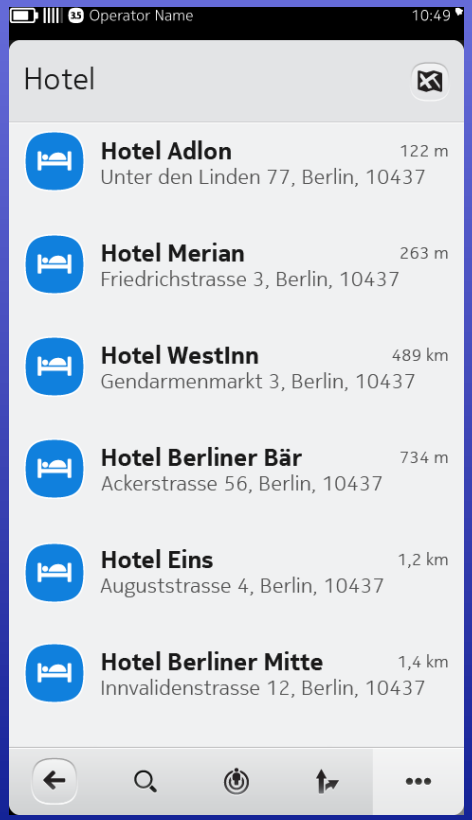
Efficient A/B Testing: *Rank Interleaving*

Version A:
Best of Eat'n'Drink



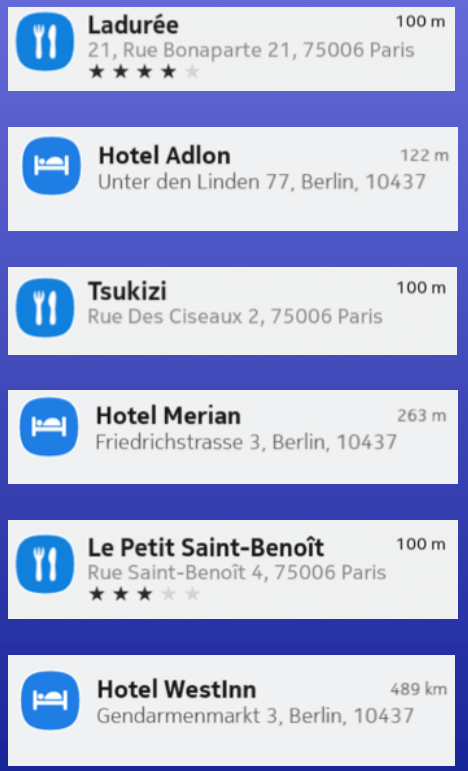
+

Version B:
Best of Hotels



=

Rank Interleaving:
Version A + B



NOKIA

Randomized Mixing of Result Lists

- Interleaved list is filled with pairs of results, one item from each version. Coin toss decides who comes first.

Version A
1. *alpha*
2. *beta*
3. *gamma*
4. *delta*
5. *epsilon*

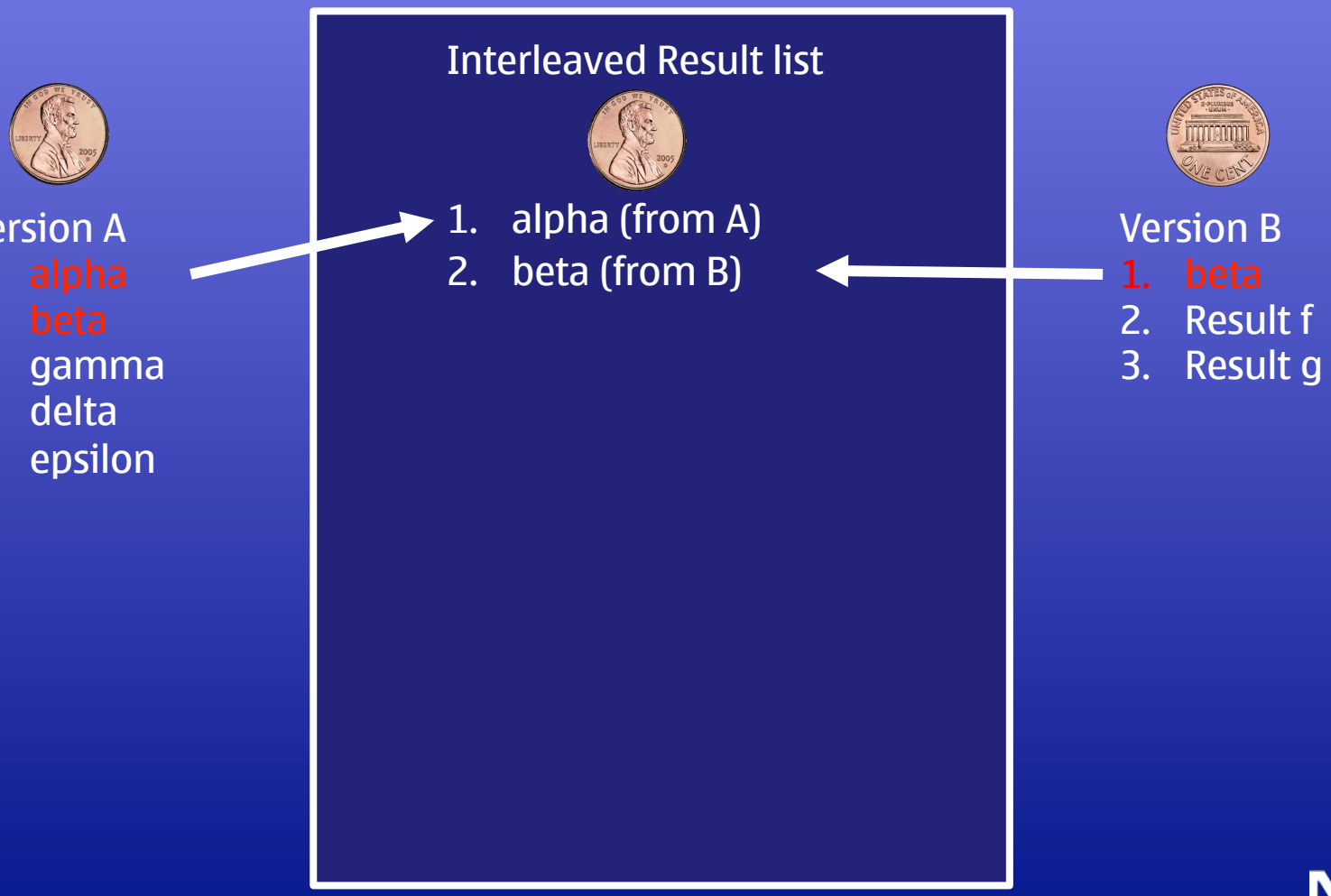
Interleaved Result list

<empty>

Version B
1. *beta*
2. *kappa*
3. *tau*

A/B Interleaving: Randomized Mixing of Lists

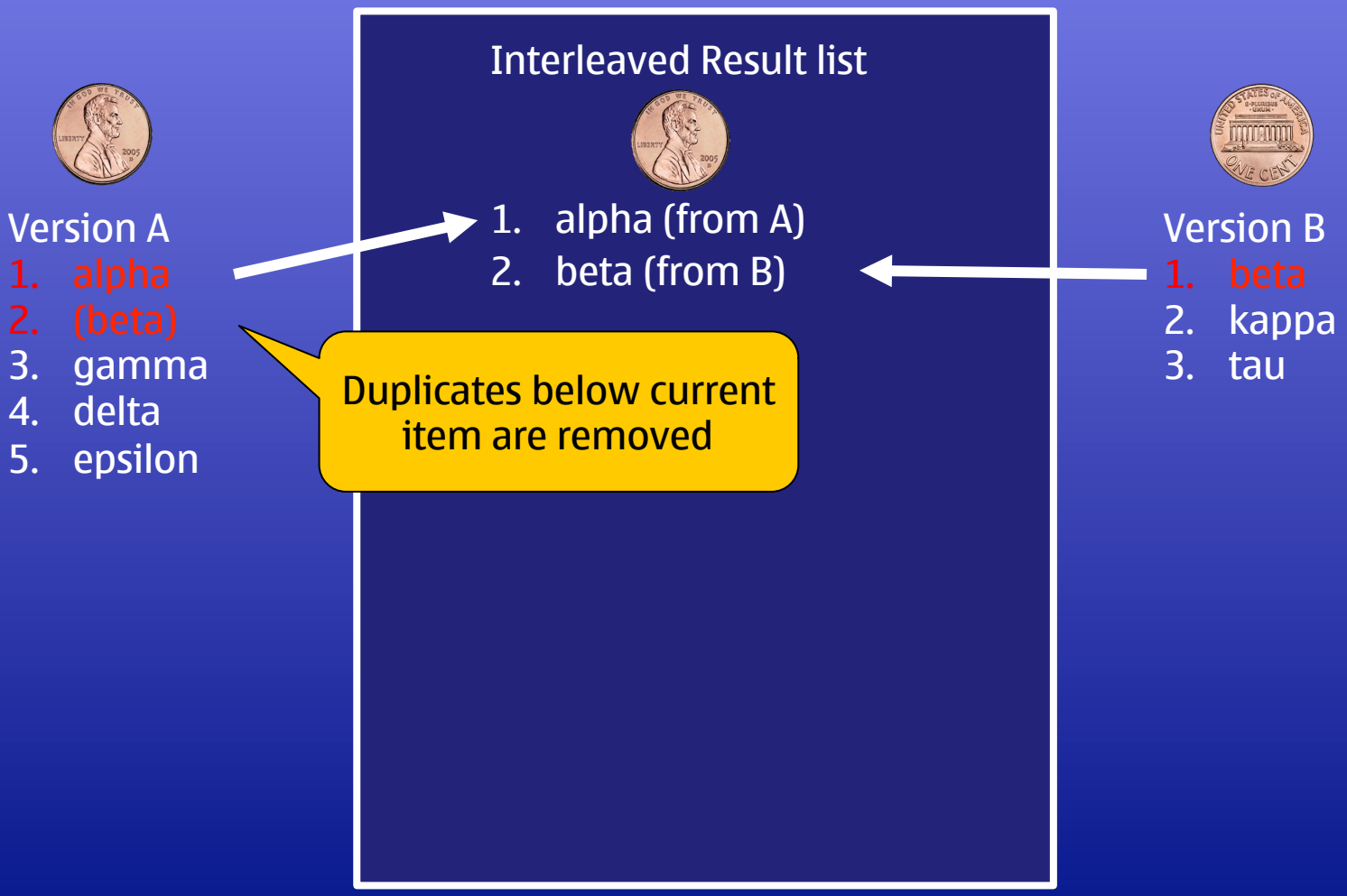
- Interleaved list is filled with pairs of results, one item from each version. Coin toss decides who comes first.



NOKIA

A/B Interleaving: Randomized Mixing of Lists

- Interleaved list is filled with pairs of results, one item from each version. Coin toss decides who comes first.



NOKIA

A/B Interleaving: Randomized Mixing of Lists

- Interleaved list is filled with pairs of results, one item from each version. Coin toss decides who comes first.

Version A
1. alpha
2. (beta)
3. gamma
4. delta
5. epsilon

Interleaved Result list

1. alpha (from A)
2. beta (from B)

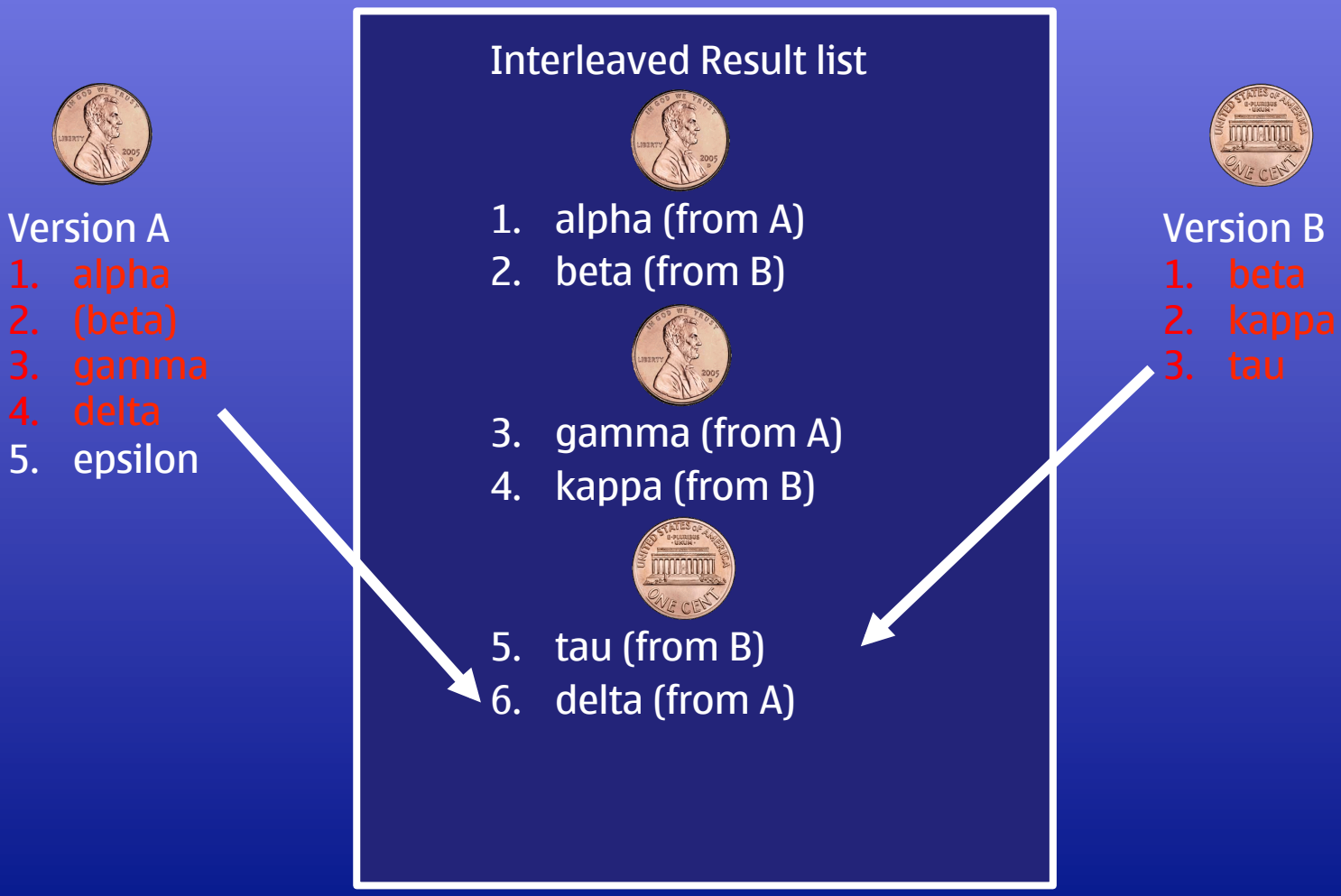
3. gamma (from A)
4. kappa (from B)

Version B
1. beta
2. kappa
3. tau

NOKIA

A/B Interleaving: Randomized Mixing of Lists

- Interleaved list is filled with pairs of results, one item from each version. Coin toss decides who comes first.



A/B Interleaving: Randomized Mixing of Lists

- Interleaved list is filled with pairs of results, one item from each version. Coin toss decides who comes first.

Version A
1. alpha
2. (beta)
3. gamma
4. delta
5. epsilon

Interleaved Result list

1. alpha (from A)
2. beta (from B)

3. gamma (from A)
4. kappa (from B)

5. tau (from B)
6. delta (from A)

7. epsilon (from A, extra)

Leftover results are appended but clicks are not counted

Version B
1. beta
2. kappa
3. tau

NOKIA

A/B Interleaving: Randomized Mixing of Lists

- Interleaved list is filled with pairs of results, one item from each version. Coin toss decides who comes first.

Version A
1. alpha
2. (beta)
3. gamma
4. delta
5. epsilon

Final list shown to user

1. alpha (from A)
2. beta (from B)
3. gamma (from A)
4. kappa (from B)
5. tau (from B)
6. delta (from A)
7. epsilon (from A, extra)

Version B
1. beta
2. kappa
3. tau

NOKIA

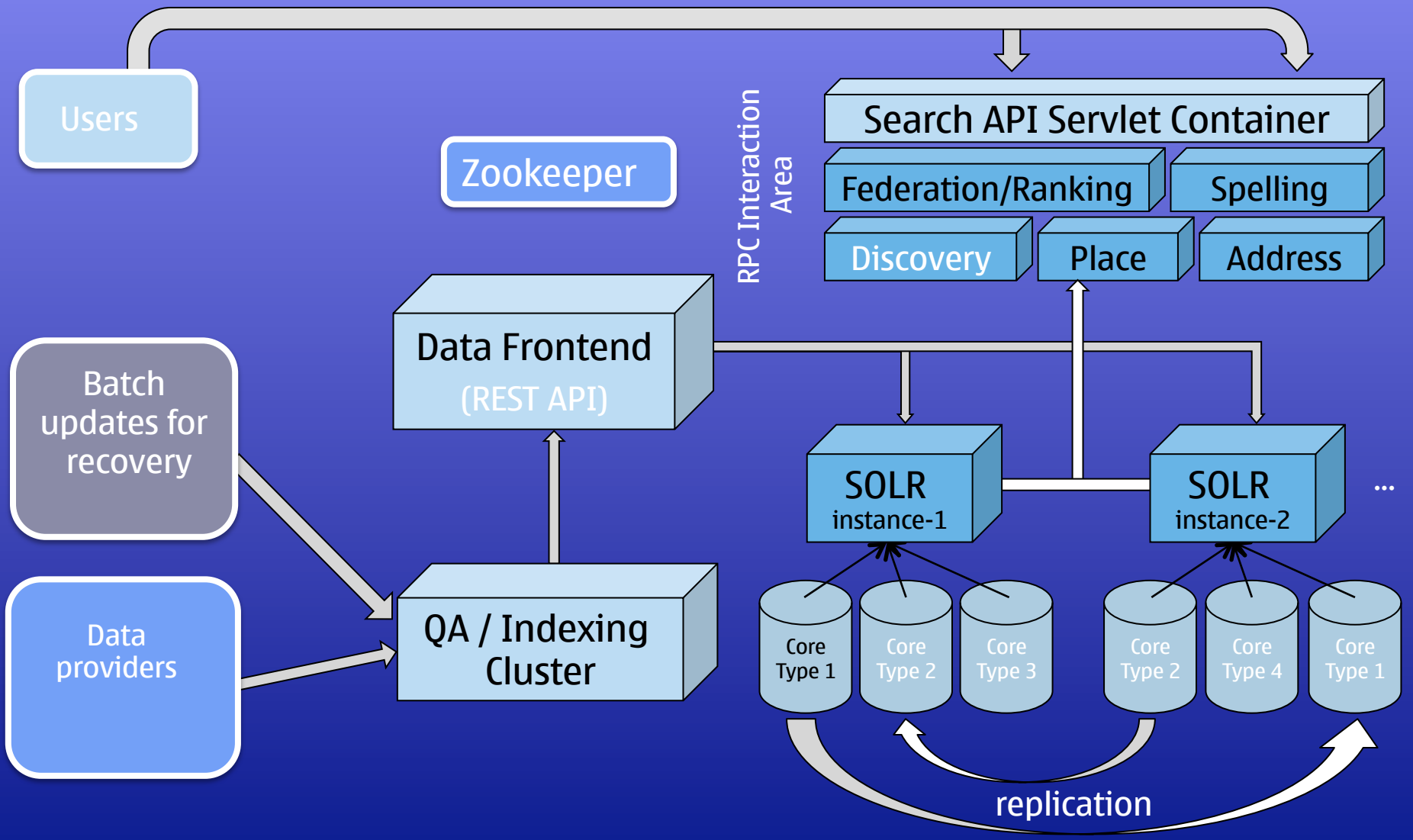
Declaring A Winner

- Statistical Significance Test
- Input (after hadoop-based log-processing...)
  - Number of clicks on version A
  - Number of clicks on version B
- G-Test:
  - improved version of Pearson's Chi-squared test.
  - G > 6.635 corresponds to 99% confidence level
- Null hypothesis:
  - Frequency of counts is equally distributed over both versions.
- Test statistic:

$$G = 2 \sum_{i \in \{A,B\}} [\text{counts i}] \ \ln\left(\frac{[\text{counts i}]}{[\text{total counts}/2]}\right)$$

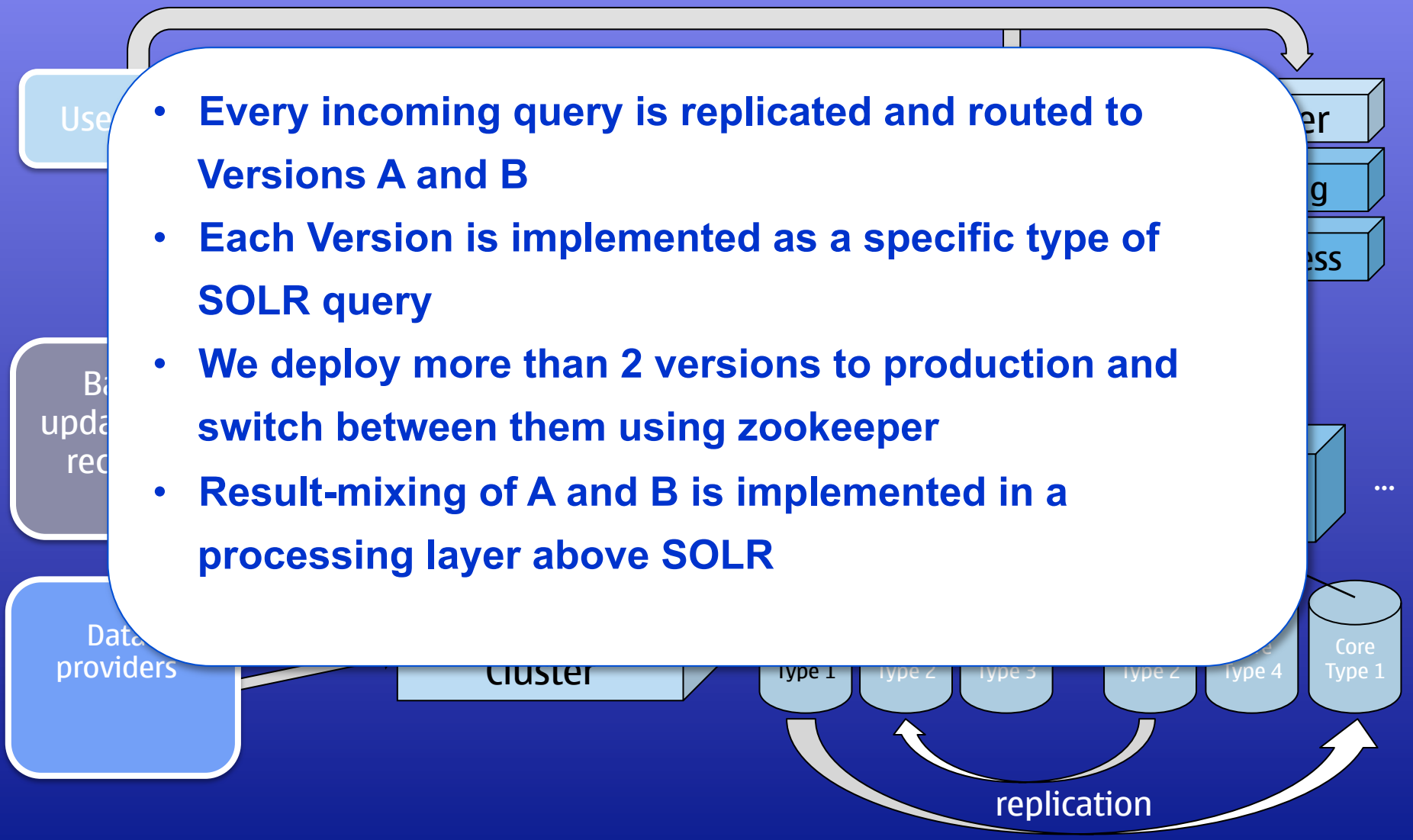
NOKIA

Managing Multiple Versions



NOKIA

Managing Multiple Versions



- Every incoming query is replicated and routed to Versions A and B
- Each Version is implemented as a specific type of SOLR query
- We deploy more than 2 versions to production and switch between them using zookeeper
- Result-mixing of A and B is implemented in a processing layer above SOLR

NOKIA

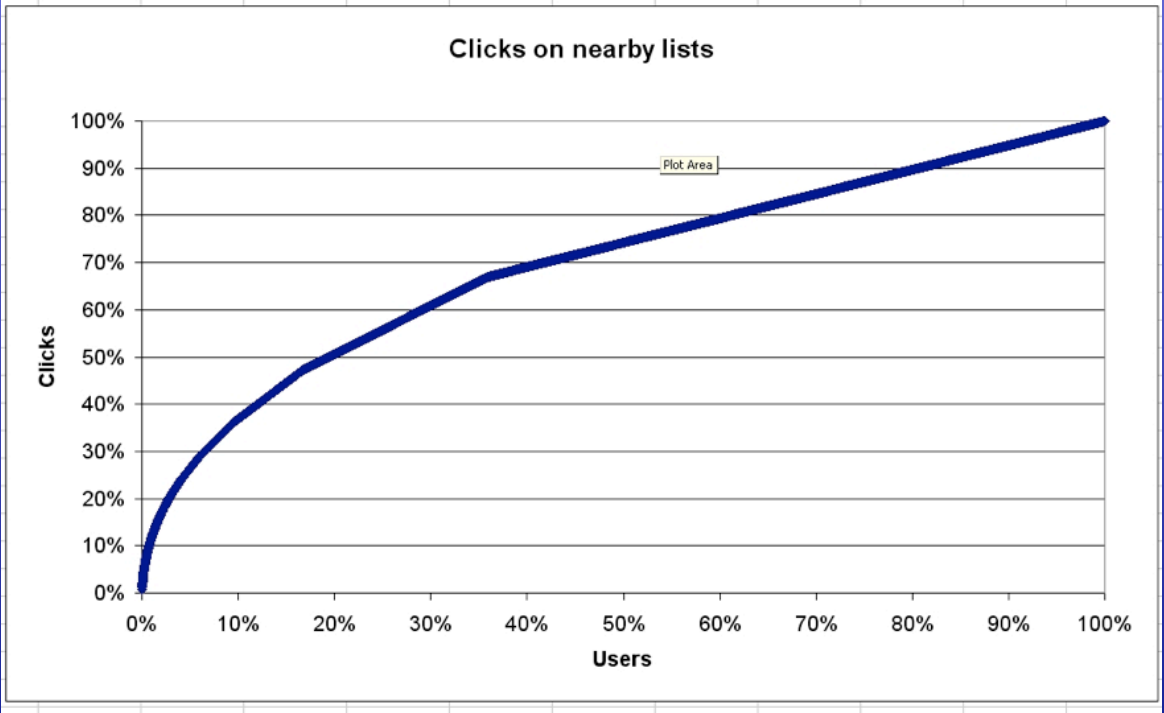
Caveat 1: Randomization

- don’t confuse users with changing results, i.e.: provide a consistent user experience
- Solution:
  - Random generator is seeded with USER-ID for each query.
  - Each user gets his personal random generator.

NOKIA

Caveat 2: Healthy Click Data

- we are relying on the integrity of transmitted user actions
- sensitive to log contamination (unidentified QA, spam)
- user-clicks plot:

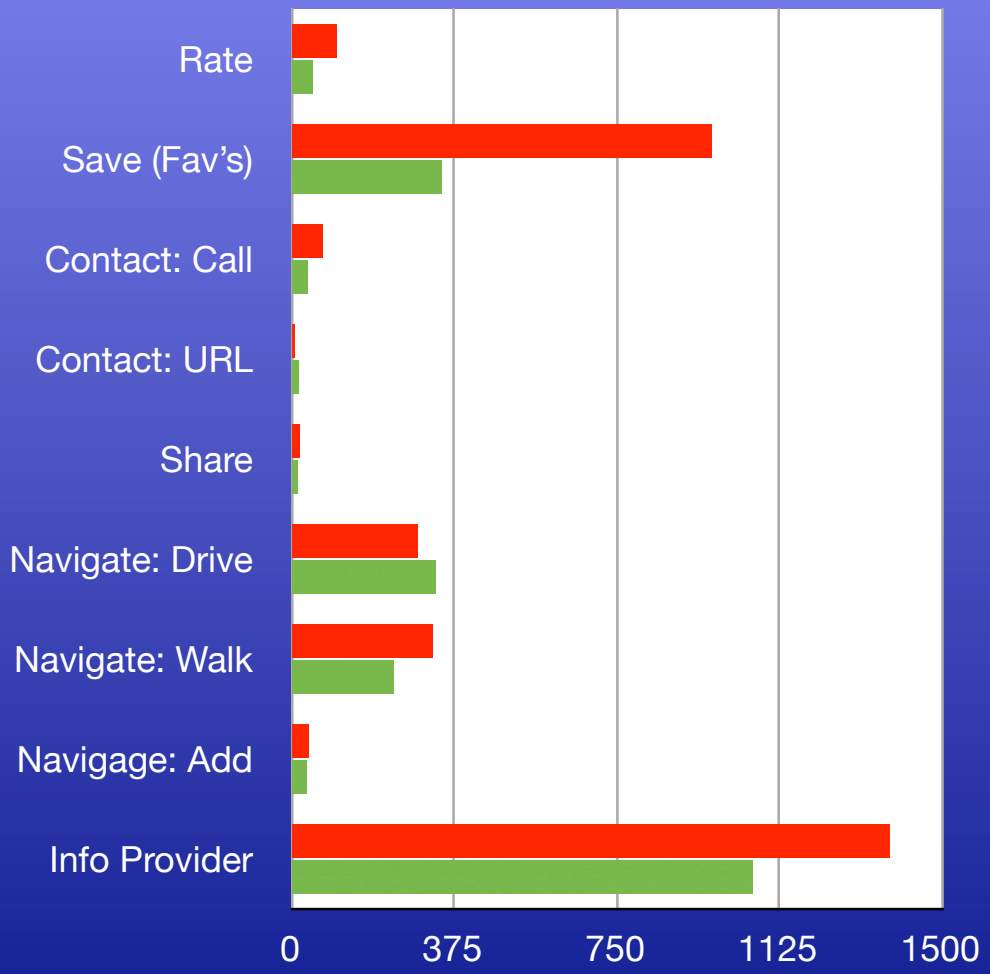


NOKIA

Caveat 3: A/B Clicks vs. Coverage

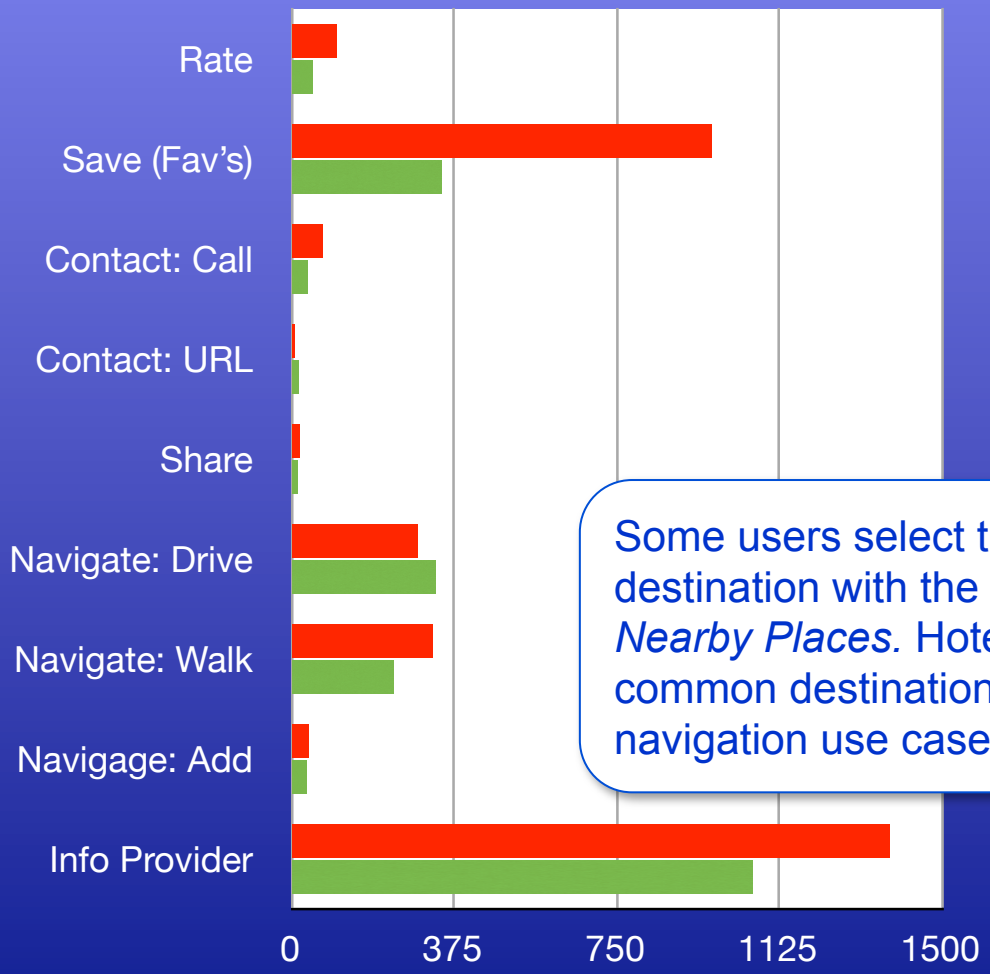
- Coverage = non-empty responses (in percent)
- For example
  - A/B interleaving of eat&drink vs. eat&drink + going out
  - difference is not significant
  - But coverage different, percentage of responses with POIs nearby:
    - 60% eat&drink
    - 62% eat&drink + going out
- Higher coverage decides in case there is no statistical difference

Case Study: Eat'n'Drink versus Hotels: Not the User Behaviour we had expected!



NOKIA

Case Study: Eat'n'Drink versus Hotels: Not the User Behaviour we had expected!



Some users select their driving destination with the help of *Nearby Places.* Hotels are a common destination in the car navigation use case.

NOKIA

Summary

- use A/B Rank Interleaving to optimize result relevance
- Rank Interleaving is easy to implement. Works.
- in a distributed search architecture manage your A/B test configurations conveniently using Zookeeper
- harness your hadoop/search analytics stack for A/B test evaluations
- don't make assumptions about your users!

- **[Joachims et al 2008]:**

"How Does Clickthrough Data Reflect Retrieval Quality?"

NOKIA

Thanks!

Get in touch: *hannes.kruppa@nokia.com*

Nokia Maps "Place Discovery" Team, Berlin:

Hannes Kruppa, Steffen Bickel, Mark Waldaukat,
Felix Weigel, Ross Turner, Peter Siemen

NOKIA