

# use cases as a bridge between benchmarking and validation

Jussi Karlgren  
starting points from the CHORUS coordination action

may 27, 2009, brussels

benchmarking is the main research method of information access research.

relevance as a target notion has served well in a constrained information and task space.

## when moving to new text and to multimedia

- the space of possible application broadens;
- the leeway the data affords for individual interpretation is larger.

how to can we extend the target notion?

when moving to new text and to multimedia

- the space of possible application broadens;
- the leeway the data affords for individual interpretation is larger.

how to can we extend the target notion?

when moving to new text and to multimedia

- the space of possible application broadens;
- the leeway the data affords for individual interpretation is larger.

how to can we extend the target notion?

when moving to new text and to multimedia

- the space of possible application broadens;
- the leeway the data affords for individual interpretation is larger.

how to can we extend the target notion?

topical text is one thing. going to multimedia will be different:

- information vs entertainment?
- push vs pull?
- lean-back vs lean-forward?
- satisfaction rather than optimisation?
- shoulder-to-shoulder vs individual vs asynchronous?



topical text is one thing. going to multimedia will be different:

- information vs entertainment?
- push vs pull?
- lean-back vs lean-forward?
- satisfaction rather than optimisation?
- shoulder-to-shoulder vs individual vs asynchronous?

topical text is one thing. going to multimedia will be different:

- information vs entertainment?
- push vs pull?
- lean-back vs lean-forward?
- satisfaction rather than optimisation?
- shoulder-to-shoulder vs individual vs asynchronous?

topical text is one thing. going to multimedia will be different:

- information vs entertainment?
- push vs pull?
- lean-back vs lean-forward?
- satisfaction rather than optimisation?
- shoulder-to-shoulder vs individual vs asynchronous?

topical text is one thing. going to multimedia will be different:

- information vs entertainment?
- push vs pull?
- lean-back vs lean-forward?
- satisfaction rather than optimisation?
- shoulder-to-shoulder vs individual vs asynchronous?

topical text is one thing. going to multimedia will be different:

- information vs entertainment?
- push vs pull?
- lean-back vs lean-forward?
- satisfaction rather than optimisation?
- shoulder-to-shoulder vs individual vs asynchronous?

what characterises success for an information access multimedia project?

system algorithm performance?

interface design issues?

or is content everything?

- topical retrieval measures? naah.
- benchmarking based on the wrong scenario? nope.

what characterises success for an information access multimedia project?

system algorithm performance?

interface design issues?

or is content everything?

- topical retrieval measures? naah.
- benchmarking based on the wrong scenario? nope.

what characterises success for an information access multimedia project?

system algorithm performance?

interface design issues?

or is content everything?

- topical retrieval measures? naah.
- benchmarking based on the wrong scenario? nope.



what characterises success for an information access multimedia project?

system algorithm performance?

interface design issues?

or is content everything?

- topical retrieval measures? naah.
- benchmarking based on the wrong scenario? nope.

what characterises success for an information access multimedia project?

system algorithm performance?

interface design issues?

or is content everything?

- topical retrieval measures? naah.
- benchmarking based on the wrong scenario? nope.

what characterises success for an information access multimedia project?

system algorithm performance?

interface design issues?

or is content everything?

- topical retrieval measures? naah.
- benchmarking based on the wrong scenario? nope.

we need more knowledge of user behaviour and task context.

central to all behavioural studies: reliability vs validity.

validation needs user studies to be true.

## lack of overlap!

multi-media projects do not stress interaction design

generalisable guidelines need interaction craft.

the CHORUS project believes use case analysis bridges this gap.

## lack of overlap!

multi-media projects do not stress interaction design

generalisable guidelines need interaction craft.

the CHORUS project believes use case analysis bridges this gap.



## lack of overlap!

multi-media projects do not stress interaction design

generalisable guidelines need interaction craft.

the CHORUS project believes use case analysis bridges this gap.

use cases can be established without empirical basis, from previous knowledge

use cases can set parameters for benchmarking evaluation

use cases must be validated by user studies, which means empirical study

use cases can be established without empirical basis, from previous knowledge

use cases can set parameters for benchmarking evaluation

use cases must be validated by user studies, which means empirical study

use cases can be established without empirical basis, from previous knowledge

use cases can set parameters for benchmarking evaluation

use cases must be validated by user studies, which means empirical study

# objective

get rid of many user studies.

get rid of much hand-waving.

it is easy to say “user behaviour”.

how can we act on leverage information from them wrt algorithms?

how can we get sustainable results from user studies?)

# objective

get rid of many user studies.

get rid of much hand-waving.

it is easy to say “user behaviour”.

how can we act on leverage information from them wrt algorithms?

how can we get sustainable results from user studies?)

## objective

get rid of many user studies.

get rid of much hand-waving.

it is easy to say “user behaviour”.

how can we act on leverage information from them wrt algorithms?

how can we get sustainable results from user studies?)

this session:

more on use cases

more on practice in projects meeting reality

more on evaluation



should we try to cut corners by agreeing on some standard use cases?