DSL/DSM evaluation

Andre Furtado
Ankica Barisic
Juha-Pekka Tolvanen
Michael Maximilien
DSL evaluation challenges

- Multiple metrics
- General purpose methods may not be adequate
- Ad hoc process
Some evaluation metrics

- Development time
- Quality of generated code
- Quality of the generated product
- Performance of generated code
- Memory usage of generated code
- Learnability
- Morale of developers
- Subjective metrics (survey)
- ...

...
Food for thought

**DSM evaluation can be domain-specific**

- What works for a medicine domain might not for games
- What is the amount of users we are able to collect data from?
- What is the size of data to collect?
- Who is the report is targeted at?
  - Management, Developers, Testers, ...?

**Data is king**

- A/B testing
- Not always feasible due to restricted target audience

**Safe bet: controlled experiments**

- Control groups
- Comparison with previous solutions

**Methods from UX can be incorporated**
Food for thought

Qualitative approaches are important to complement data
  • Where users are getting stuck?
  • What part of the language syntax is not clear?

We should be careful to not compare apples with oranges
  • “Is this DSL better than Java?”

The comparison should not focus on the language alone
  • Toolset, IDE integration
  • Documentations/how to’s (videos, tutorials, guidelines)

Evaluation is a discipline on itself
  • Encourage evaluation experts/teams to emerge
Food for thought

Evaluation should not be an afterthought

- Know who your user is
- Define your metrics/quality attributes

Release early and often

Call for action

- Can we (DSM community) document a set of common evaluation standards, as in a catalogue?
- Sharing best practices on common evaluation challenges
- And also evaluation patterns
DSL/DSM evaluation

Thanks!

Andre Furtado
Ankica Barisic
Juha-Pekka Tolvanen
Michael Maximilien