

AIR Breakout Group

Evaluation of Adaptive Systems

AIR 2008

Outline

- What to evaluate
 - Why
- According to what criteria
- With what measures
- Using what methods
- Requiring what resources

What to Evaluate

- Why adaptation at all?
 - Take account of context, in order to enhance user experience, e.g. engagement, pleasure, usability? effectiveness? To account for differences in situations.
- What do we mean by *context*?
 - State of user, past, present... future? static and dynamic; goals, knowledge of topic, system, expectations ...
 - Environment - surroundings, location, time,
 - System
- How “well” the system adapts to context
- The entire information-seeking episode, as well as parts of it

According to what Criteria

- Information-oriented
 - Task/goal achievement (i.e effectiveness)
 - Usefulness w.r.t. e.g. parts of episode, whole episode
 - How well the system learns
 - Relevance
- Interaction-oriented
 - How well the system learns
 - Usability, learnability, pleasurability

With what Measures?

- Time to ... e.g. task completion
- Success in ... e.g. task completion
- Searcher judgments of usefulness ...
- Time taken for system to learn contextual features
- Mean accuracy in contextual feature identification

With what Methods

- User studies?
 - Longitudinal
 - Observational
 - Experimental
- Log analysis?
- Combination of two?
- Other?

With what Resources?

- Problems
 - Replicability
 - Cumulativity (building on previous work)
 - Cost
 - Amount of data
- Solutions?
 - Common resources
 - Common methods