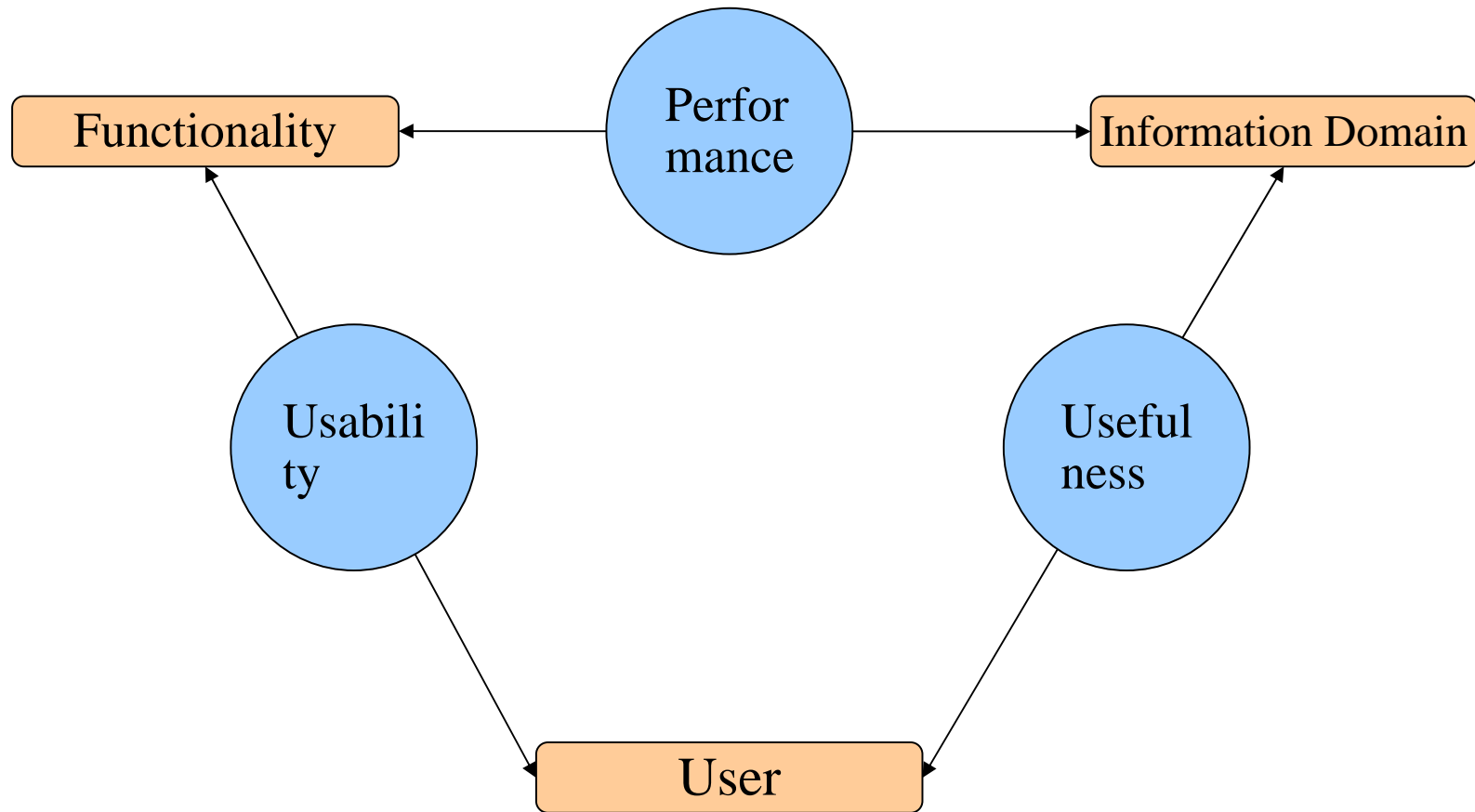# Evaluation of DL Systems
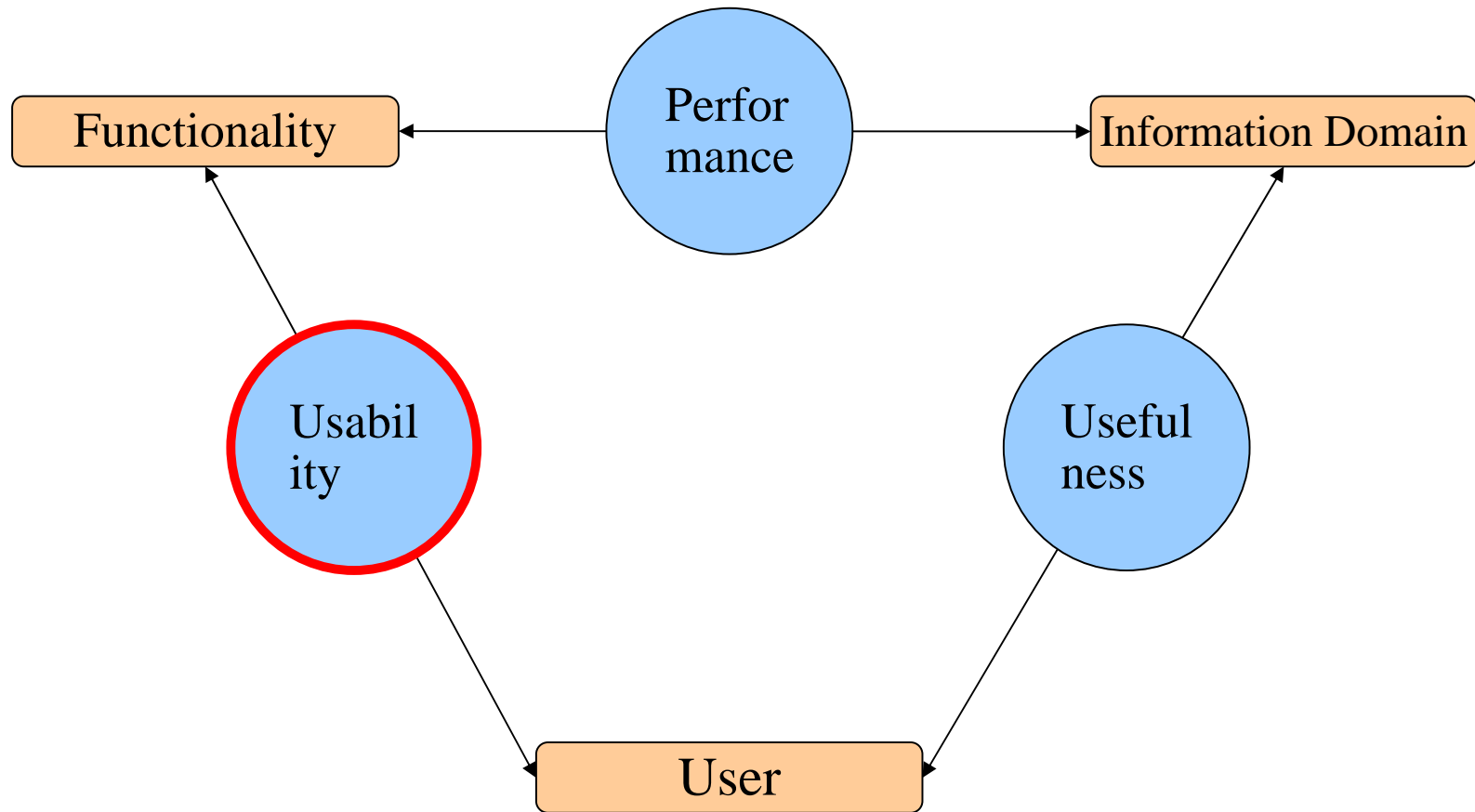
Norbert Fuhr

University of Duisburg

Saracevic's evaluation framework

- ## Construct: WHAT to evaluate
  (aspects/components of the DL)
- ## Context: WHY
  (scenarios, actors, objectives, goals/approaches, perspectives)
- ## Criteria: HOW
  (parameters, factors, measures)
- ## Methodology: HOW
  (procedures/protocols for data gathering + analysis)

The Interaction Triptych Model
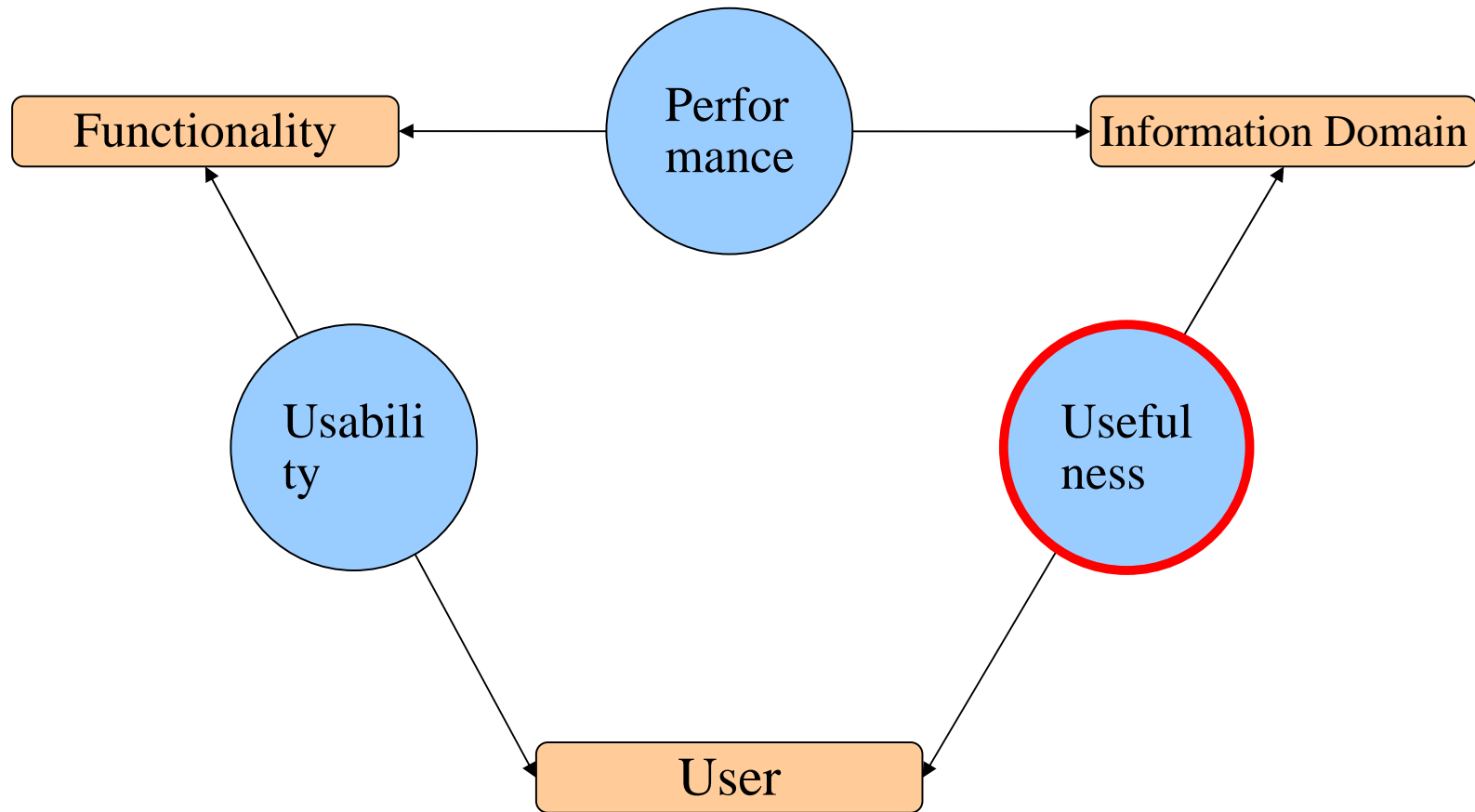
The Interaction Triptych Model

Usability (User-Functionality)

Usability evaluation techniques:

- Automated techniques
- Empirical or research-based techniques
- Comparative techniques
- Interviews, surveys, focus groups, questionnaires
- Analysis techniques

The Interaction Triptych Model

Usefulness evaluation techniques:

- User studies
- Information behaviour
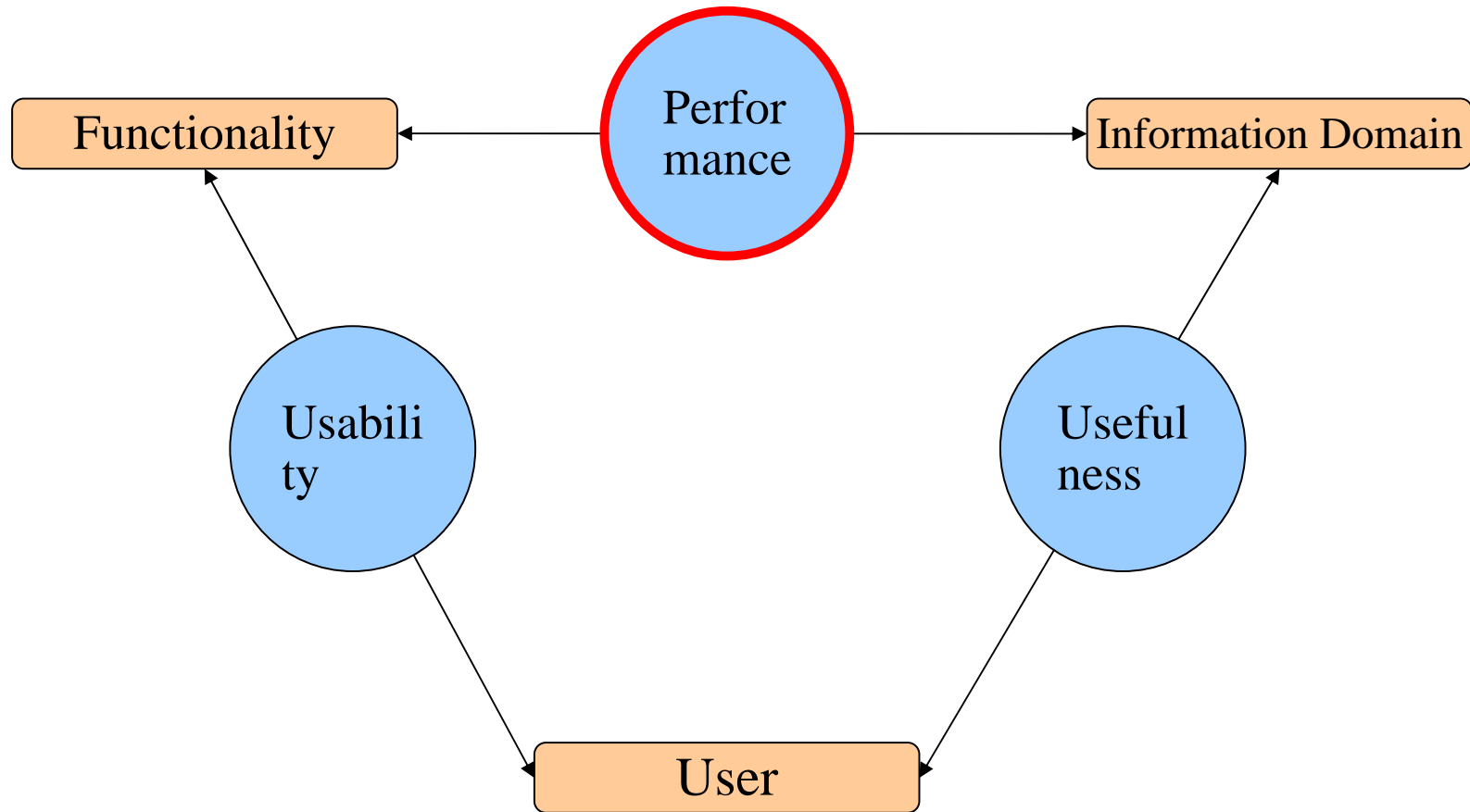- Studies related to the information domain

User studies

Evaluation criteria:

- Types of users
- Information needs
- Satisfaction factors
- Types of information handling strategies
- Tasks and task procedures
- Representation of work domains and environment
- Collaboration between users and user groups

Information Domain-related studies

- Content nature:
  - Form
  - Language
  - Method of creation
  - Type (white/grey)
- Content structure:
  - Level
  - Size
  - Cohesion
  - Delivery mode
- Administration
  - Collection Growth
  - Rights
  - License mode

The Interaction Triptych Model

Performance evaluation techniques:

- Off-line evaluation
  (laboratory setting, like in TREC, CLEF, INEX...)

- On-line evaluation
  (distributed access/computing of "live" systems)

Performance criteria

- ## Efficiency
  - response time
  - memory/CPU usage, #IOs
  - ....

- ## Effectiveness
  - recall
  - precision
  - cost-based measures
  - ....

Comparing our work with Saracevic's framework

- ## Construct: WHAT
  - → DL Reference Model + Interaction Triptych provide good basis for describing construct

- ## Context: WHY
  - → application-specific

- ## Criteria: HOW
  - → some standard criteria and metrics

- ## Methodology: HOW
  - → some standard methods

Conclusion

- DL evaluation: What? Why? How?
- Evaluation focuses on relationships between components:
  - Usability: User - Functionality
  - Usefulness: User - Information Domain
  - Performance: System - Information Domain
- Broad variety of methods for evaluating these relationships
- Lack of standard criteria and metrics