

Evaluation: Adequacy and Diagnostic Evaluation

COMP61332: Text Mining

Week 5

Riza Batista-Navarro

Adequacy Evaluation

Evaluation as seen by **users**

- Can I carry out some task effectively and productively
- Context of use is critical

Direct comparison of systems difficult

Judging **external quality**: characteristics are interdependent

- sometimes one hurts another
- sometimes one helps another

the degree to which a system is free from faults in its specification, design and implementation

how easy it is to learn and use a system

use of system resources

a mean time between failures

prevention of unauthorized or improper access to a program and data

a system can be used, without modification, in new applications

degree to which a system is free from error; how well it does the job

degree to which a system continues to function in the presence of invalid inputs

Adaptability

Integrity

Efficiency

Robustness

Correctness

Reliability

Usability

Accuracy

the degree to which a system is free from faults in its specification, design and implementation

how easy it is to learn and use a system

use of system resources

a mean time between failures

prevention of unauthorized or improper access to a program and data

a system can be used, without modification, in new applications

degree to which a system is free from error; how well it does the job

degree to which a system continues to function in the presence of invalid inputs

Adaptability

Integrity

Efficiency

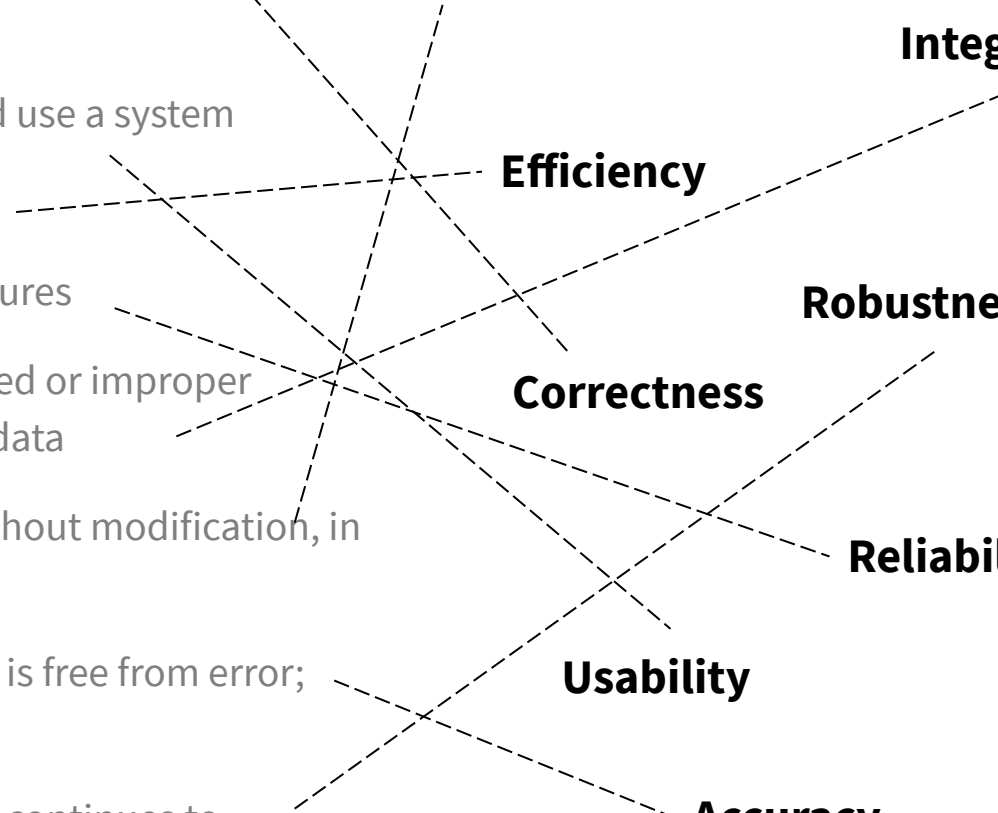
Robustness

Correctness

Reliability

Usability

Accuracy



Trade-offs between characteristics

	Cor.	Usa.	Eff.	Reli.	Inte.	Adap.	Acc.	Rob.
Correctness			↑	↑			↑	↓
Usability						↑	↑	
Efficiency	↓			↓	↓	↓	↓	
Reliability	↑				↑		↑	↓
Integrity			↓	↑				
Adaptability					↓			↑
Accuracy	↑		↓	↑		↓		↓
Robustness	↓	↑	↓	↓	↓	↑	↓	

Diagnostic Evaluation

Evaluation as seen by developers

Internal quality

the ease with which you can modify a software system

Portability

the extent you can modify a system for uses other than originally specified

Reusability

the extent you can modify a system to operate in a different environment

Maintainability

the extent to which you can use parts of a system in other systems

Testability

the ease with which you can read and understand the source code of a system

Understandability

Flexibility

the degree to which you can unit-test and system-test a system

Readability

the ease with which you comprehend a system (at a more general level)

the ease with which you can modify a software system

the extent you can modify a system for uses other than originally specified

the extent you can modify a system to operate in a different environment

the extent to which you can use parts of a system in other systems

the ease with which you can read and understand the source code of a system

the degree to which you can unit-test and system-test a system

the ease with which you comprehend a system (at a more general level)

Portability

Reusability

Maintainability

Testability

Understandability

Flexibility

Readability

Diagnostic Evaluation

Test suites

- For an NLP-based system, design and coverage of diagnostic test suites is hard
- Range of phenomena system expected to tackle is hard to anticipate

May emphasise idealised linguistic aspect, may be of reduced use in face of ungrammatical (real) text