



TECHNISCHE  
UNIVERSITÄT  
WIEN  
VIENNA  
UNIVERSITY OF  
TECHNOLOGY

# Towards an analytical evaluation of preservation strategies

Presentation for the ERPANET Workshop

By Carl Rauch and Andreas Rauber

10<sup>th</sup>-11<sup>th</sup> of Mai 2004, Vienna

Department for Software Technology & Interactive Systems  
Vienna University of Technology

- 
- We have
    - collections with different file formats and preservation requirements
  - We have
    - myriads of potential preservation approaches (various converters, emulators, metadata schemes,...)
  - We need
    - a way to decide which one to pick rather than un-transparent „out-of-the-guts“ decisions
-

---

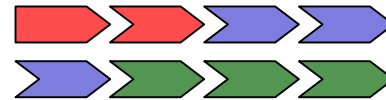
- Introduction

- Utility Analysis

-  Set objectives

-  Evaluate alternatives

-  Define preferences and decide



- Summary

---

# Selecting a preservation strategy

---

Problem → Requirements → Solutions

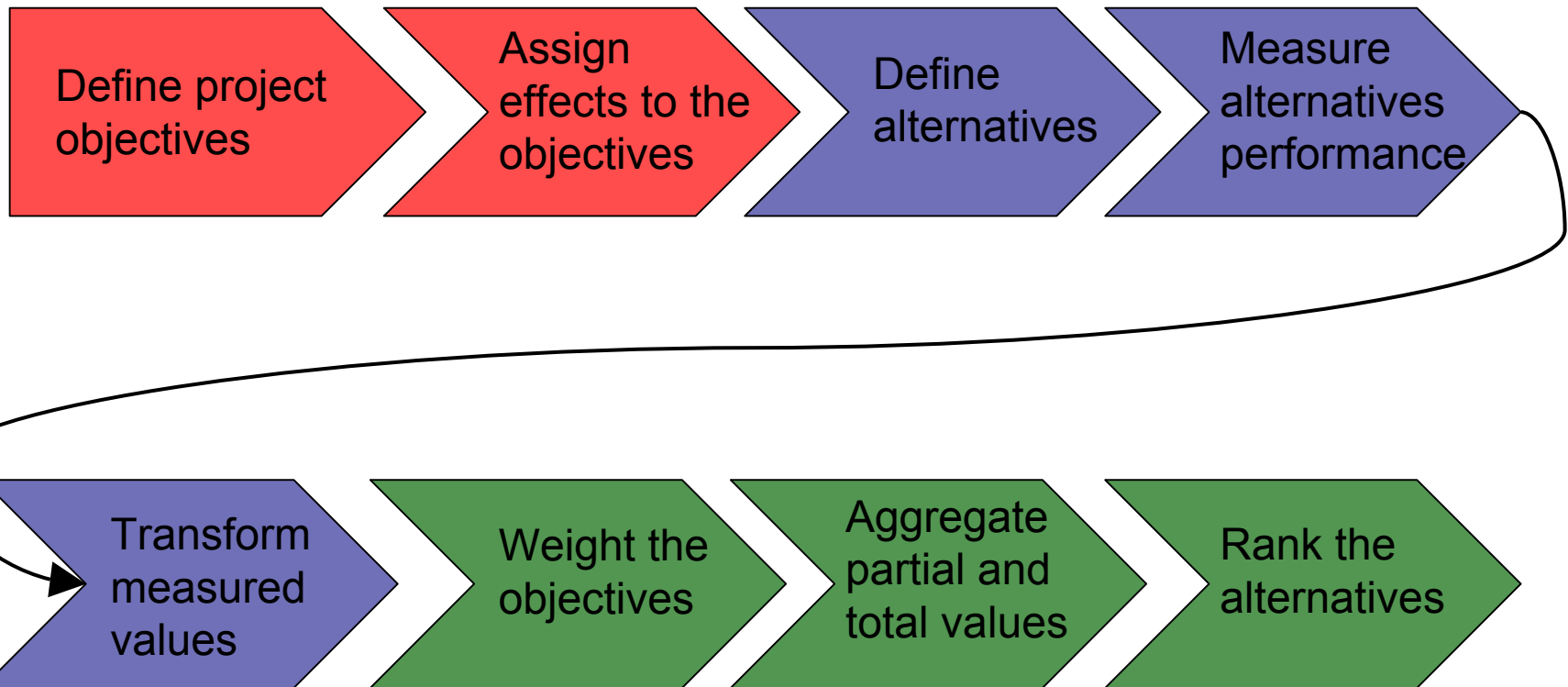
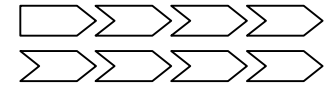
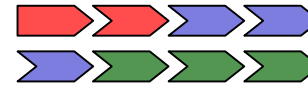
- Several different preservation strategies, where no single one excels the others in all circumstances
- Different requirements for different file collections
- Steady change and development of strategies and tools

- Strategies that obey very different requirements
- Means to make strategies comparable
- Measures to be equally applicable to new preservation strategies

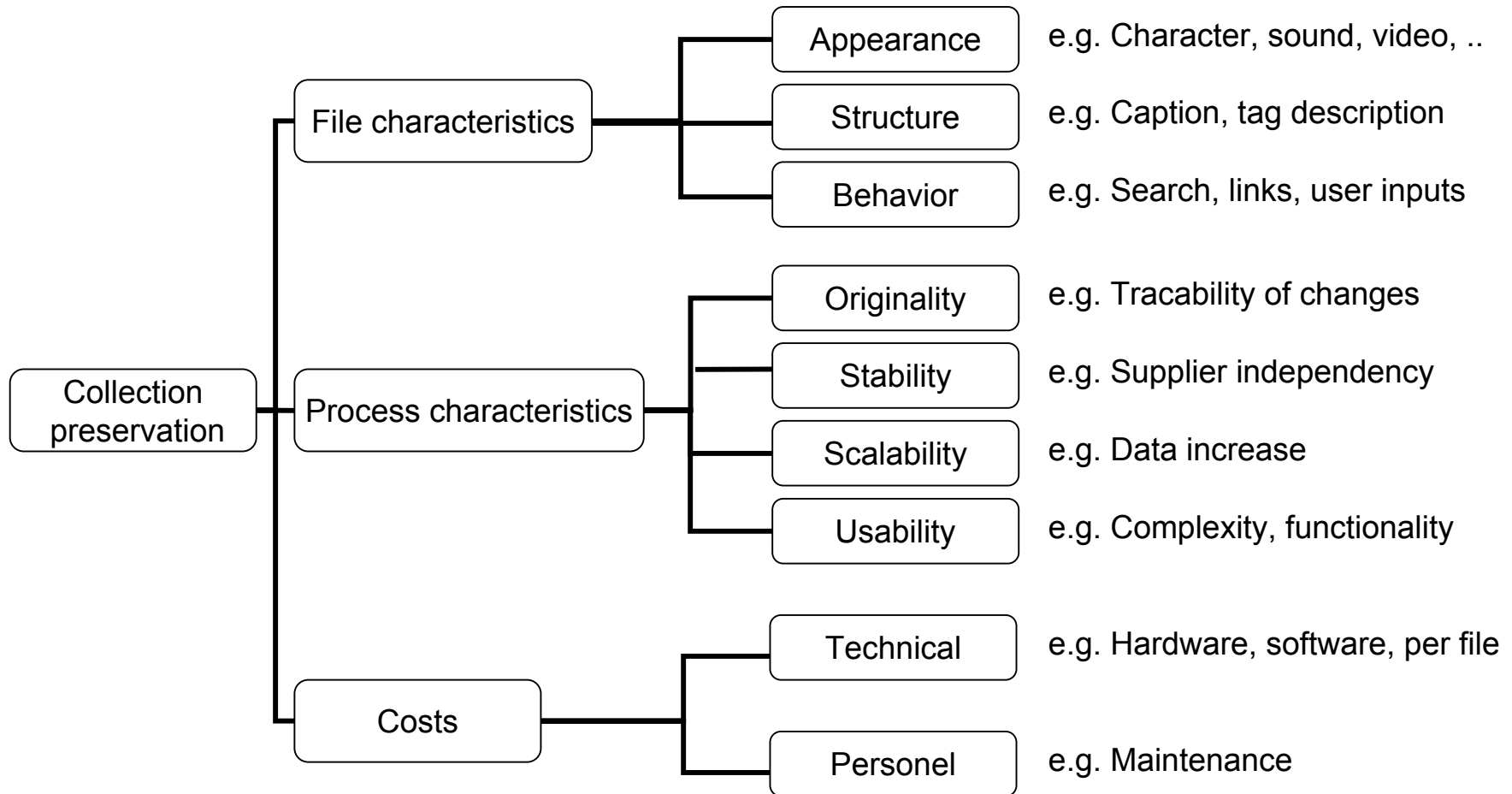
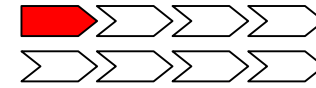
- Generic framework, which can be easily applied to specific environments
- Decision support system, which clearly ranks possible preservation solutions

- 
- Developed in the 1970s
  - Applied mainly for infrastructure projects, such as dams, bridges, neighbourhoods
  - Well expandable
  - Adapted to fit the preservation requirements
-

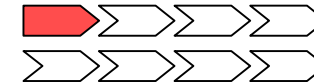
# Utility Analysis procedure



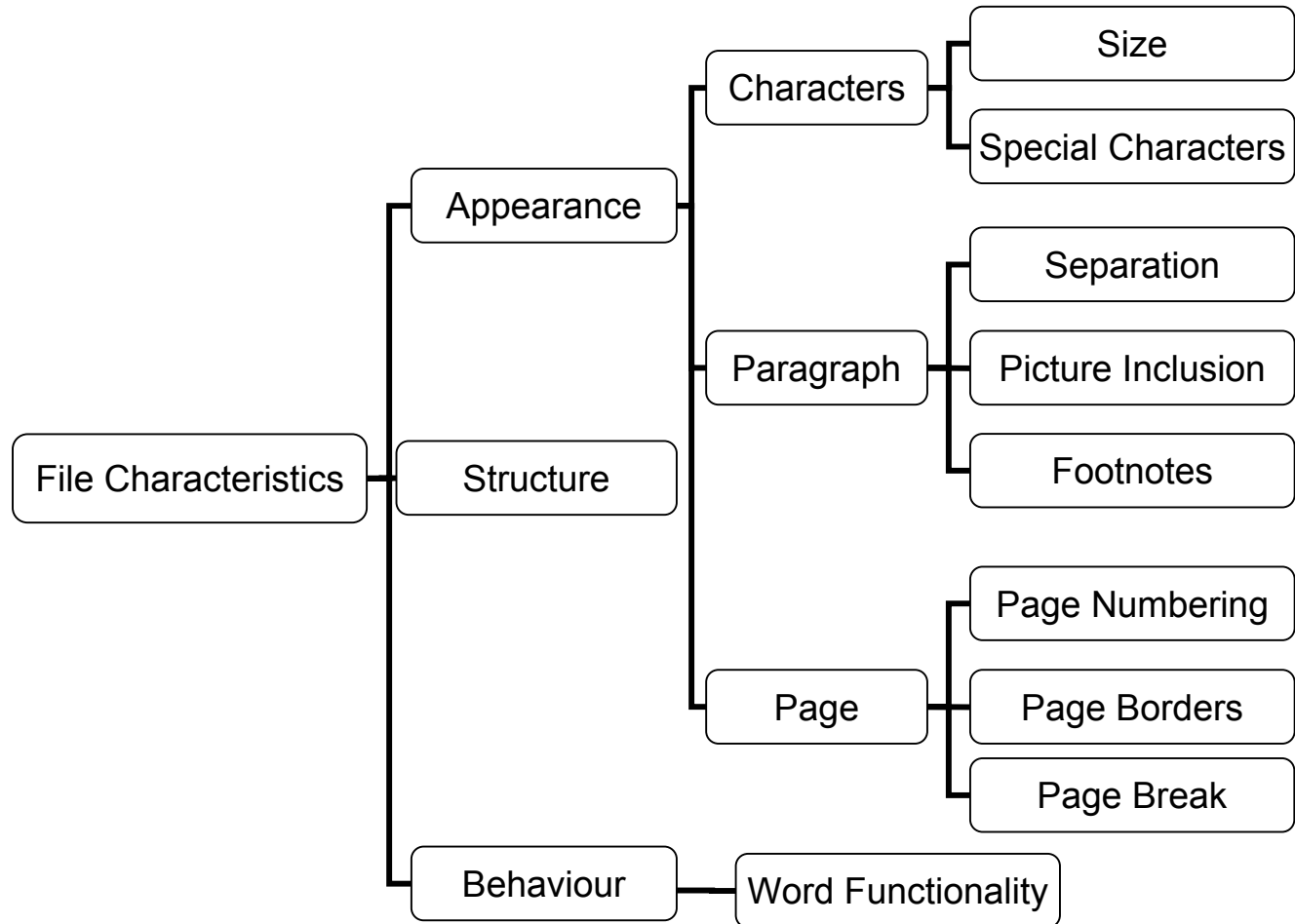
# Define project objectives



# Implemented objective tree

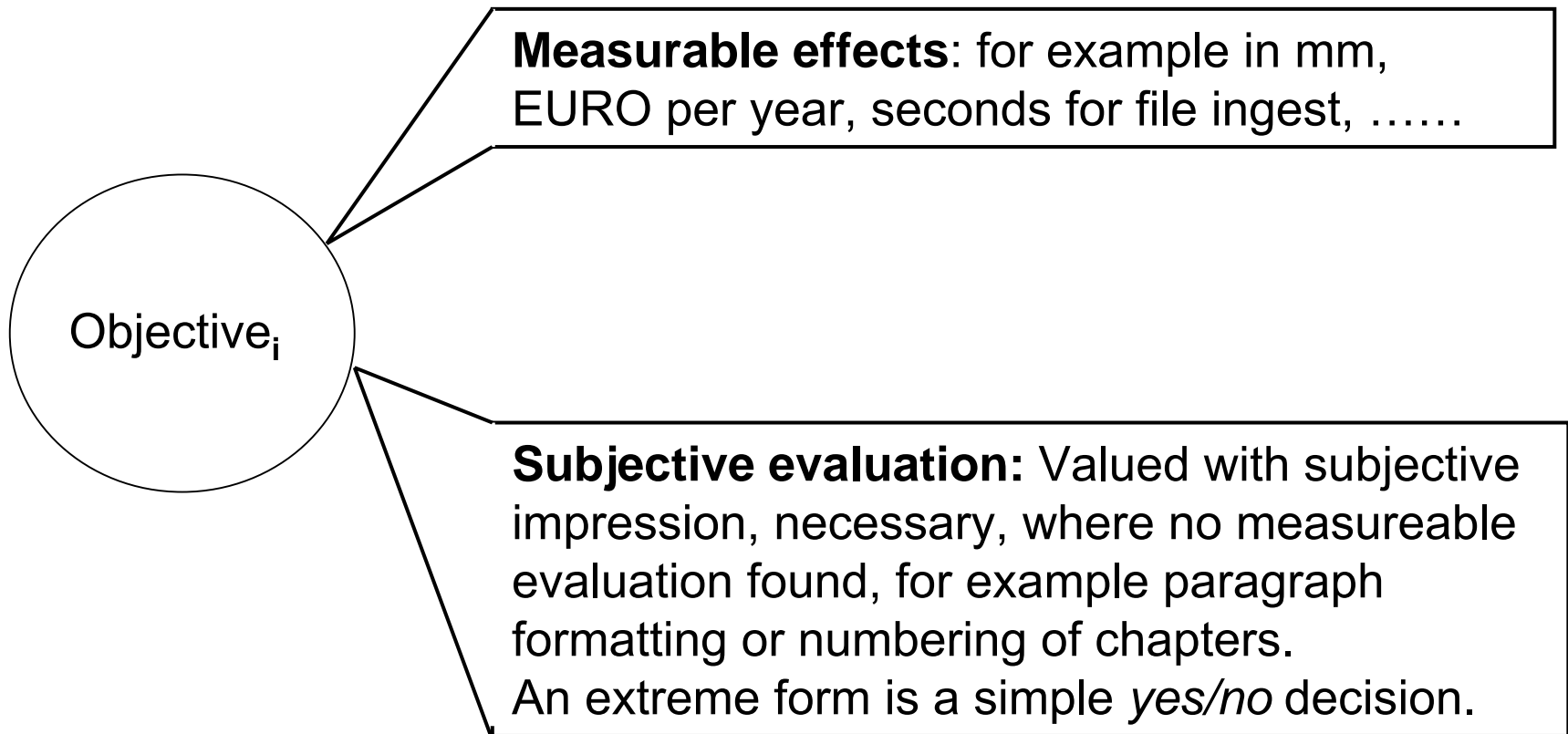
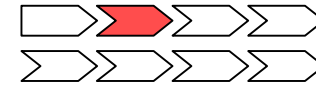


High Aspect	Medium Aspect	Low Aspect	Medium Aspect

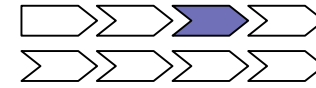


# Assign effects to objectives

---



# Definition of alternatives



## ■ Migration & Standardisation

- *Migrate documents to Adobe PDF*
- *Migrate documents to OpenOffice.org*
- *Migrate documents to PostScript*
- *Migrate documents to a newer version of MS Word*

## ■ Emulation & Encapsulation

- *Encapsulate digital objects*

## ■ Computer Museum

- *Try to preserve the hardware environment*

## ■ Digital Tablet

- *Try to construct a digital tablet*

## ■ No change to the strategy

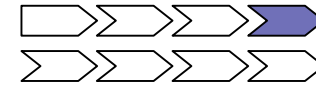
- *Do not adapt the strategy*

## ■ No preservation effort

- *Do not take care of preservation*

# Alternatives' evaluation

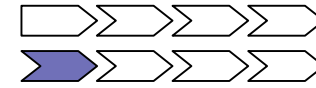
---



- Measure of the alternatives' performance, using either:
  - Original files
  - Files from a testbed

	Newer MS Word version	OpenOffice.org Writer	PDF 5.0	No changes at all
Page borders	0 mm	+ 3 mm	0 mm	0 mm
Ingest: sec. per file	10 sec	10 sec	15 sec	0 sec
Software costs per year	50 €	0 €	0 €	0 €
Numbering of chapters	3	N.A.	5	5
Paragraph formatting	4	2	5	5

# Transform measured values



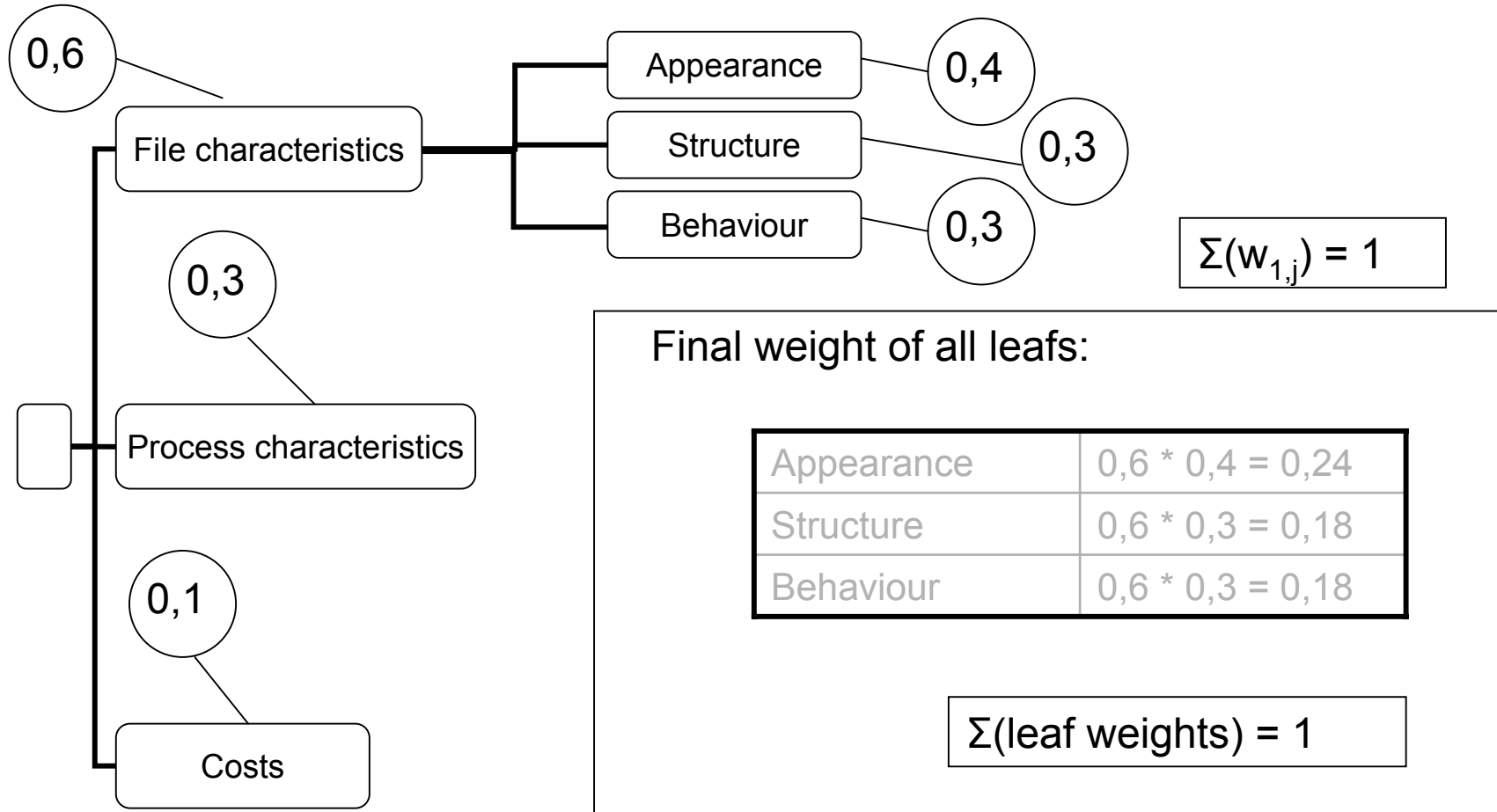
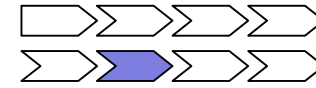
## ■ Define the transformation table:

	5	4	3	2	1	N.A
Page borders	+/- 0 mm	+/- 1 mm	+/- 2 mm	+/- 3 mm	+/- 4 mm	> 4mm
Ingest: sec. per file	0 -5 sec	5-10 sec	10-15 sec	15-25 sec	25-40 sec	>.40 sec
Software costs per year	0 €	1-30 €	31-50 €	51-70 €	71-100 €	> 100 €
Numbering of chapters	1	2	3	4	5	N.A.
Paragraph formatting	1	2	3	4	5	N.A.

## ■ Transform the results to make them comparable

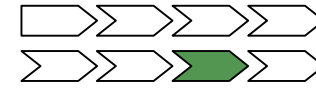
	Newer MS Word version	OpenOffice.org Writer	PDF 5.0	No changes at all
Page borders	5	2	5	5
Ingest: sec. per file	4	4	3	5
Software costs per year	3	5	5	5
Numbering of chapters	3	N.A.	5	5
Paragraph formatting	4	2	5	5

# Weighting



# Aggregating part values

---



- Part values per objective

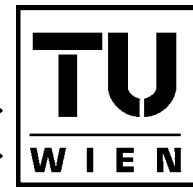
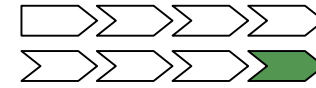
Leaf Weights x Transformed Values

- Total value per alternative

Sum of all part values of a strategy  
Includes also „not acceptable“ alternatives

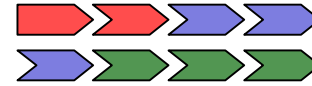
# Final Ranking

---



Ranking of the alternatives according to their total values,  
not acceptable alternatives are ranked worst

- Final sensitivity analysis, concerning non measurable influences on the decision, such as
  - expertise in a specific alternative
  - good relation to a supplier
  - ....



- 
- Composition of objective trees depend strongly on the collection's requirements
  - Different solutions vary mainly in the objective tree composition and the objective's weights
  - A few „standard“ objective trees may evolve for specific scenarios
  - We now have:
    - A powerful tool to make accountable preservation decisions
    - Decision process is transparent
-

# Next steps

---

- 
- Building and evaluation various objective trees for different preservation settings
  - Specifically, create „exhaustive“ listing of file format characteristics
  - Development of a user interface for the objective definition
  - Building a decision support system
-