Using Utility Analysis to Evaluate and Compare Preservation Strategies

Carl Rauch, Andreas Rauber

Vienna University of Technology

http://www.ifs.tuwien.ac.at
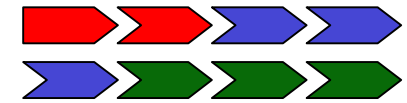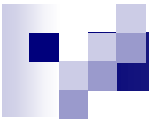
Motivation

- ## We have
  - □ collections with different file formats and preservation requirements
  - □ myriads of potential preserveration approaches (various converters, emulators, metadata schemes,…)

- ## We need
  - □ a *structured approach* to selecting and evaluating preservation solutions, rather than un-transparent „out-of-the-guts" decisions

Outline

- **Introduction**

- **Utility Analysis**
  - Set objectives
  - Evaluate alternatives
  - Define preferences and decide

- **Summary**

Selecting a preservation strategy

Problem ➡ Requirements ➡ Solution

Problem	Requirements	Solution
■ Several preservation strategies, none excels in all circumstances	■ Strategies that suit different requirements	■ Generic framework, which can be easily applied to specific environments
■ Different requirements for different collections	■ Means to make strategies comparable	
■ Steady change and development of strategies and tools	■ Measures to be equally applicable to new preservation strategies	■ Decision support system, which clearly ranks possible preservation solutions
	■ Structured approach	

Utility Analysis

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

Utility Analysis Procedure

Define project objectives → Assign effects to the objectives → Define alternatives → Measure alternatives performance →

Transform measured values → Weight the objectives → Aggregate partial and total values → Rank the alternatives
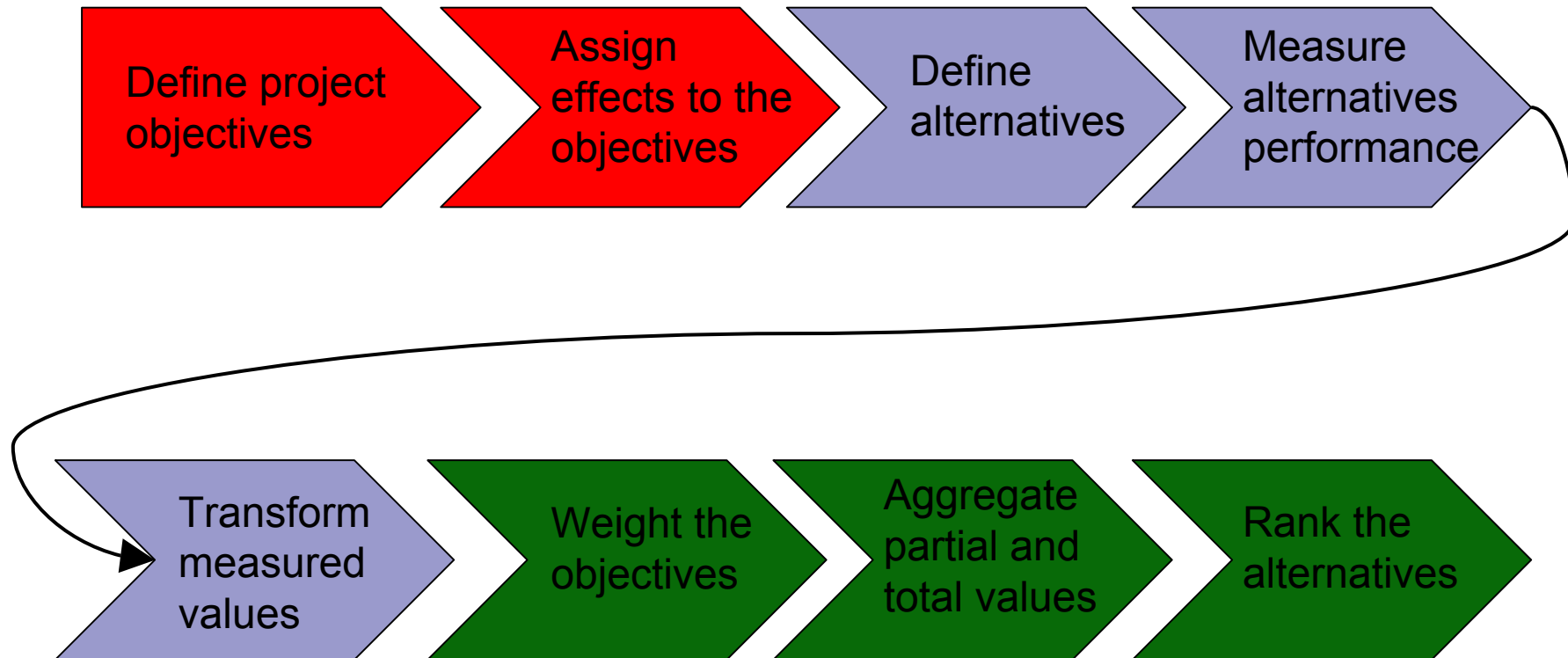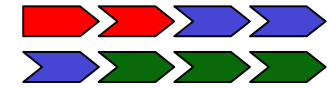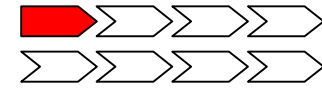
Define Project Objectives
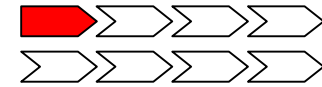
- Collect set of project objectives
- Include all requirements and desiderata
- Rather complex, extensive
- Procedure:
  - Bottom-up approach: brainstorming session
  - Top-down approach: according to generic objective tree
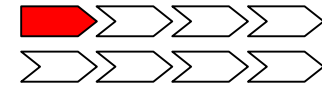 - Structure as an *Objective Tree*
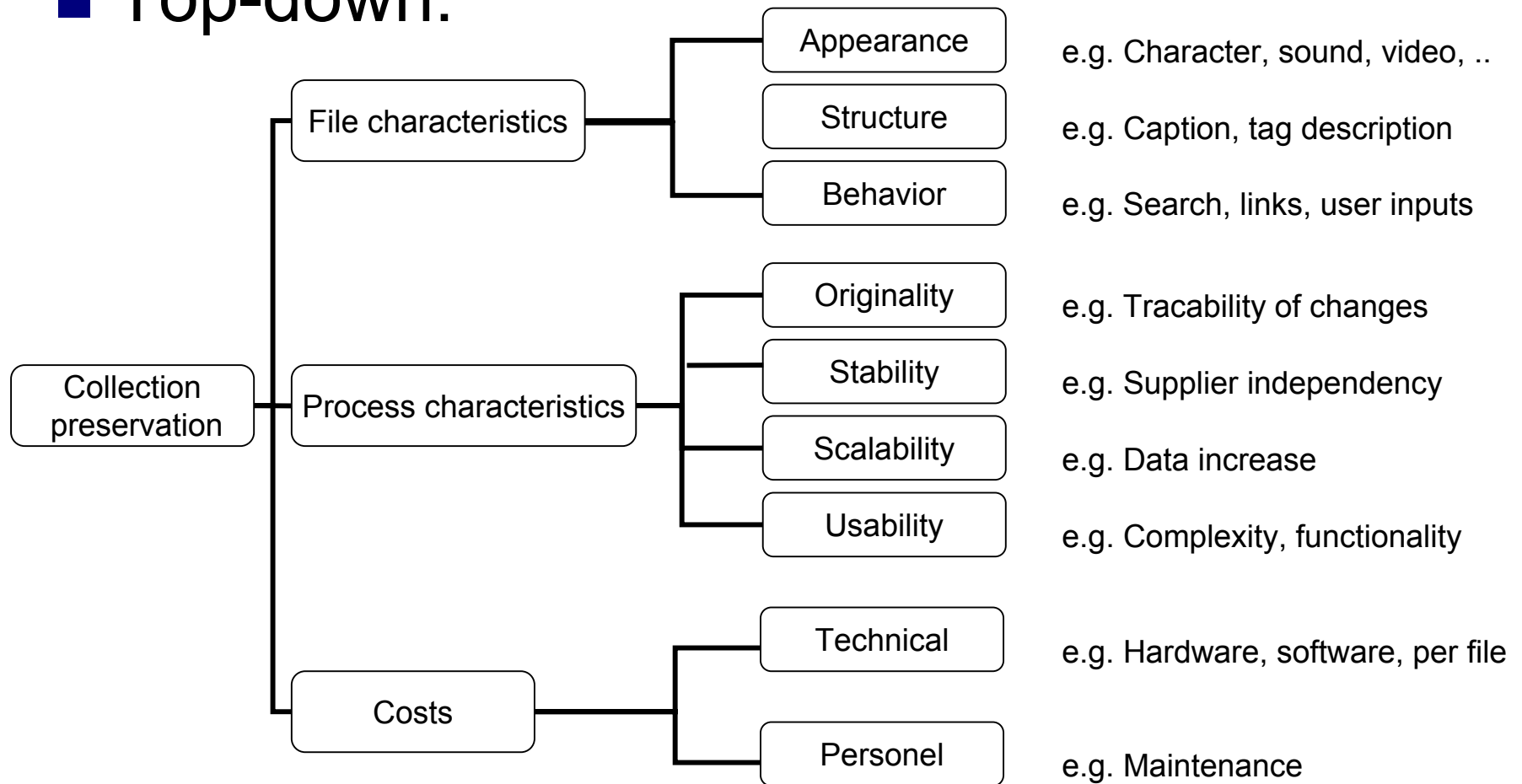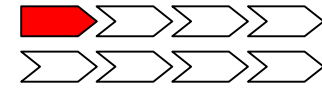
Define Project Objectives

- Bootom-up:

Define Project Objectives

- **Top-down:**

		Appearance	e.g. Character, sound, video, ..
	File characteristics	Structure	e.g. Caption, tag description
		Behavior	e.g. Search, links, user inputs
Collection preservation	Process characteristics	Originality	e.g. Tracability of changes
		Stability	e.g. Supplier independency
		Scalability	e.g. Data increase
		Usability	e.g. Complexity, functionality
	Costs	Technical	e.g. Hardware, software, per file
		Personel	e.g. Maintenance

Define Project Objectives

- Integrate:

```
File Characteristics
├── Appearance
│       ├── Characters
│       │       ├── Size
│       │       └── Special Characters
│       └── Paragraph
│               ├── Separation
│               ├── Picture Inclusion
│               └── Footnotes
├── Structure
│       └── Page
│               ├── Page Numbering
│               ├── Page Borders
│               └── Page Break
└── Behaviour
        └── Word Functionality
```
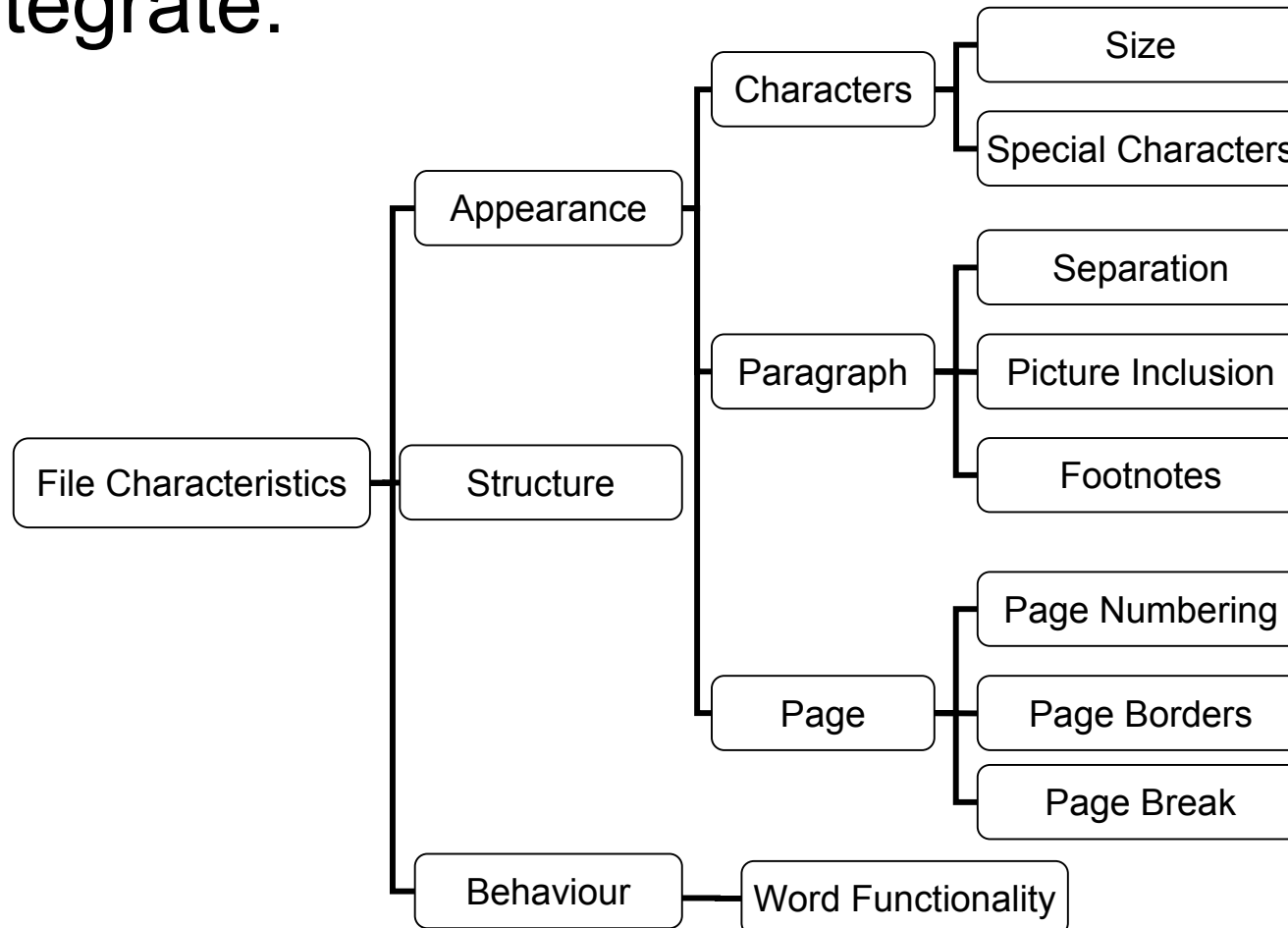
Assign effects to objectives

Objective$_i$

Measurable effects: for example in

- dpi resolution, mm difference, true/false

- EURO per year, person months

- seconds per file

- ……

Subjective evaluation:

subjective impression when no measureable evaluation possible, for example

- display quality
- look-and-feel
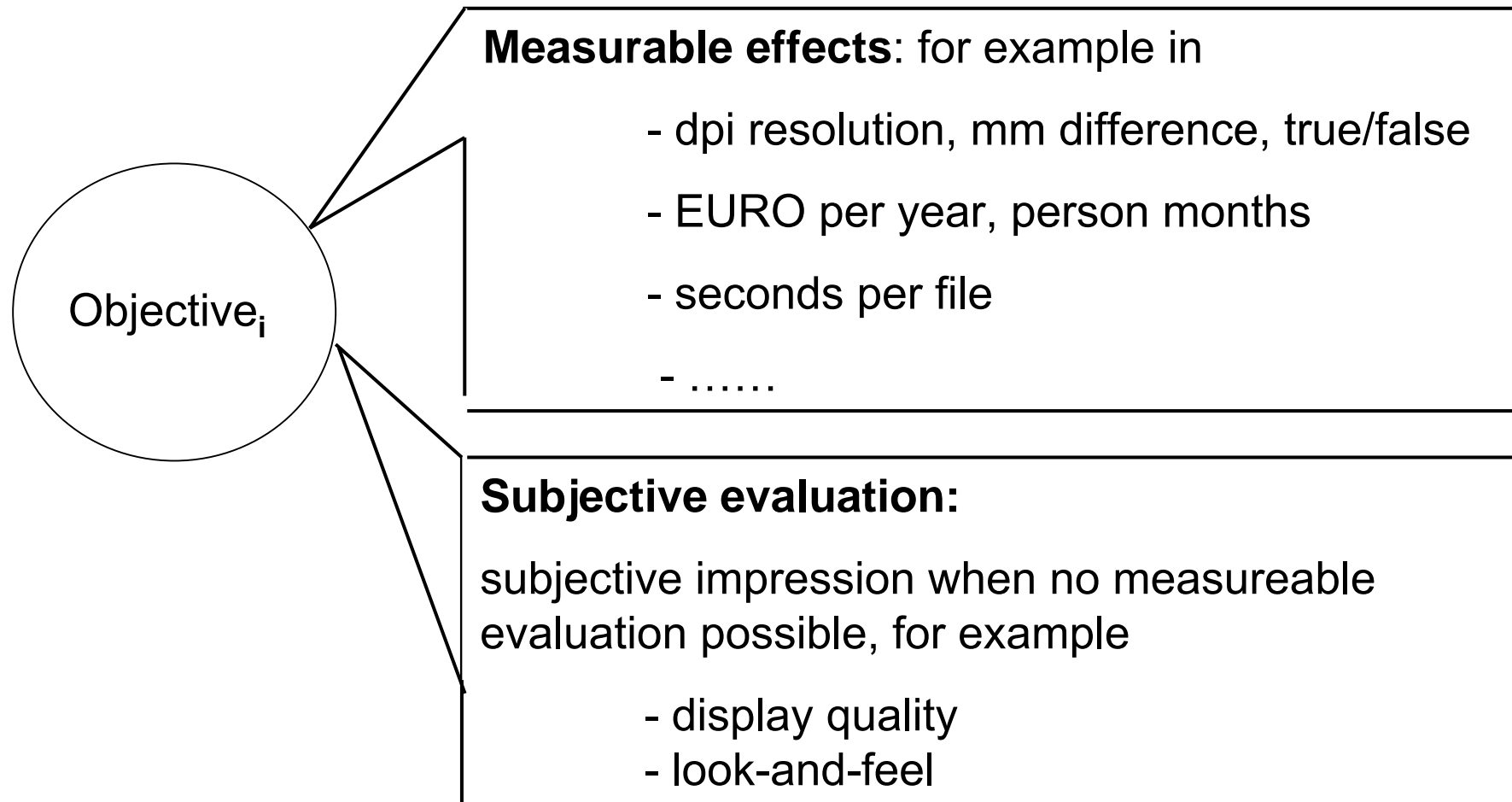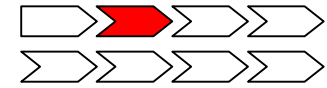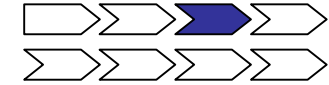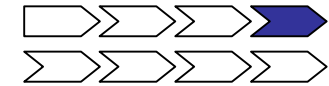
Listing Alternatives

- **Migration and Standardisation**
  - ☐ Migrate documents to Adobe PDF using XXX
  - ☐ Migrate documents to OpenOffice 1.0
  - ☐ Migrate documents to PostScript using XXX
  - ☐ Migrate documents to MS Word 2003
- **Encapsulation**
- **Hardware Museum**
- **…**
- **Maintain current strategy**
- **No action**

Alternatives' evaluation

- **Select files for evaluation**
  - ☐ Original files from collection
  - ☐ Files from a testbed
- **Ensure that they cover collection characteristics**
- **Perform preservation steps according to list of alternatives**
- **Measure results**

Alternatives' evaluation
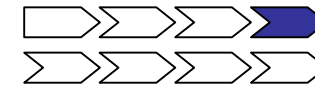
- Result:
Table of performance measures

	Word 2003	OpenOffice	PDF 5.0	No changes
Page margins	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	3	2	5	5
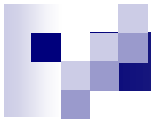
Transform Measured Values

- Need to make measured values comparable
- Define transformation table

	5	4	3	2	1	N.A
Page margins	+/- 0mm	+/- 1mm	+/- 2mm	+/- 3mm	+/- 4mm	> 4mm
Ingest: sec. per file	0 -5 sec	5-10 sec	10-15sec	15-25sec	25-40sec	>.40sec
SW costs/year	0 €	1-30 €	31-50 €	51-70 €	71-100 €	> 100 €
Chapter numbering	5	4	3	2	1	N.A.
Paragraph formatting	5	4	3	2	1	N.A.
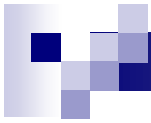
Transform measured Values

- Transform measures:

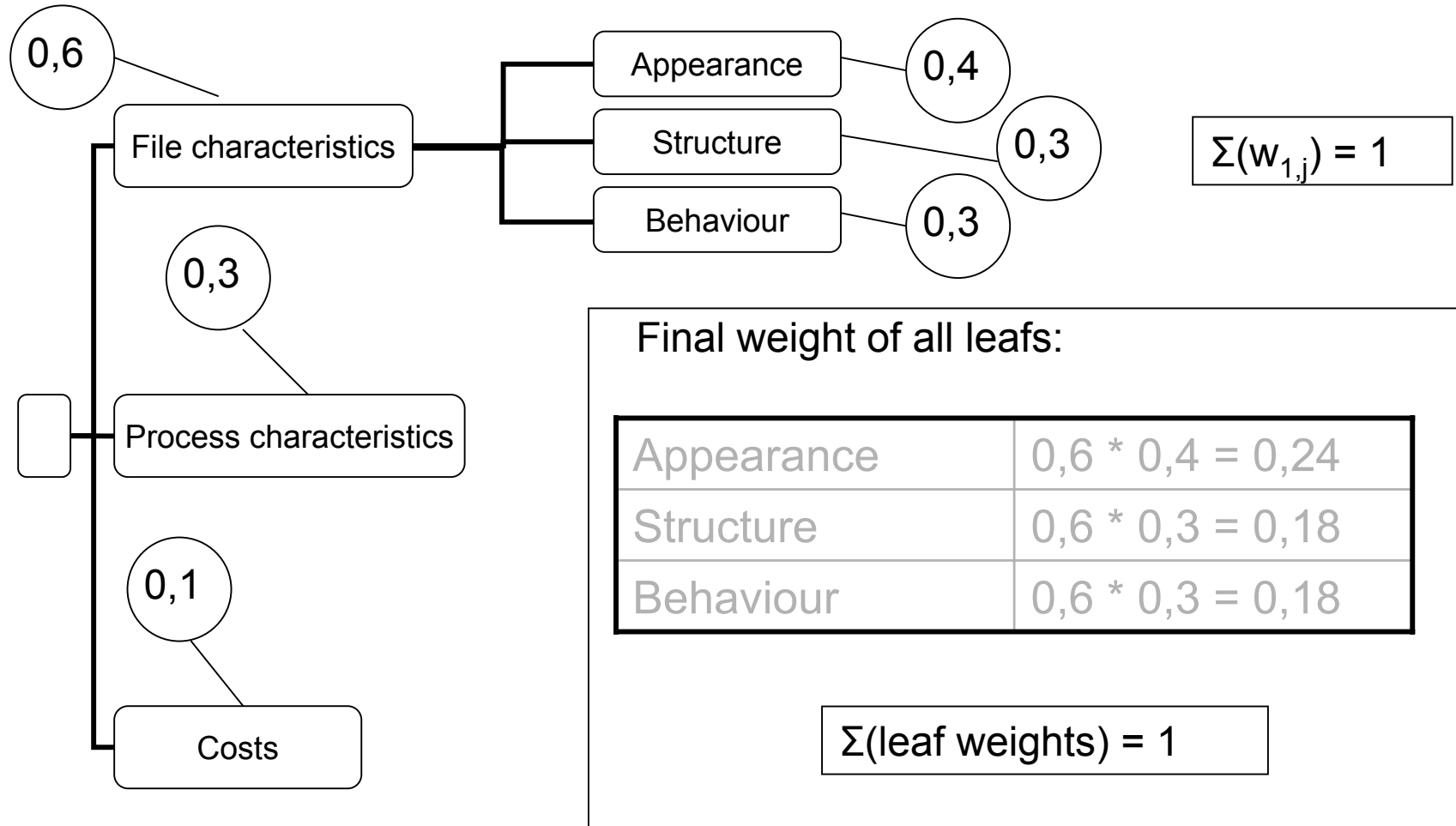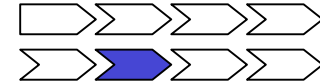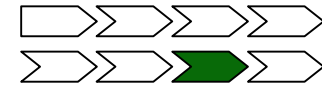|	Word 2003	OpenOffice	PDF 5.0	No changes
Page borders	5	2	5	5
Ingest: sec. per file	4	4	3	5
Software costs per year	3	5	5	5
Numbering of chapters	3	N.A.	5	5
Paragraph formatting	3	2	5	5

Weighting
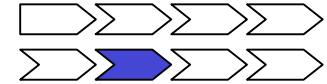
- Objectives differ in importance / priority
- Assign weights to objectives
- Basically possible right after definition of Objective Tree
- Recommended to perform after measurement and transformation
- Weights per branch level sum up to 1

Weighting

```
0,6 — File characteristics ┬ Appearance — 0,4
                           ├ Structure — 0,3
                           └ Behaviour — 0,3

0,3 — Process characteristics

0,1 — Costs
```

$$\Sigma(w_{1,j}) = 1$$

Final weight of all leafs:

| Appearance	0,6 * 0,4 = 0,24
Structure	0,6 * 0,3 = 0,18
Behaviour	0,6 * 0,3 = 0,18

$$\Sigma(\text{leaf weights}) = 1$$

Aggregating part values

- Calculate leaf values by multiplying transformed measurements with weights

- Aggregate values per alternative

- If necessary, average or min/max over different demo-files

- Provides performance per alternative according to different branch levels, i.e. objective granularities
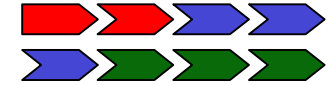
Final Ranking
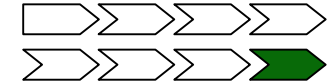
- Ranking of alternatives
- Not-acceptable alternatives are kept in ranking
- Final sensitivity analysis regarding non measurable influences on the decision, such as:
    - expertise in a specific alternative
    - good relation to a supplier
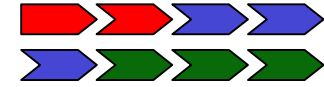   - …

Summary

- **Composition of Objective Tree depends strongly on collection requirements**

- **Different solutions vary mainly in**
  - ☐ Objective tree composition
  - ☐ Objective's weights

- **A few „standard" Objective Trees may evolve**

- **We now have:**
  - ☐ A structured approach to make accountable preservation decisions
  - ☐ A transparent decision process

Next steps

- Cooperating with institutions to elaborate "standard" Objective Trees

- Cooperate on generating "exhaustive" listings of file format characteristics

- Develop tool support for calculating different weighting scenarios

- Evolve into decision support system