A Taxonomy of Automated Typesetting Systems

Joachim Schrod  Christopher A. Rowley  Christine Detig

EuroTeX 2005 Pont-à-Mousson, France

Vision
- Formal model of document formatting
- Emphasis on text typesetting
- Investigate and document, what is automated typesetting?
- Restore automated (high-quality) typesetting to be a vigorous topic of research and development

Why
- Counterpoint to the hype of hypertext
- There is more than online publishing
- Literature is old: hence no description of current systems
- Previous work is empirical: no systematic generalisation

Strategy
What
- First steps—work in progress!
- Informal but precise description of existing document formats
- Focus on text typesetting
- Describe capabilities of formatters

What not
- Multimedia, multi-channel publishing
- Word processing
- Authoring environments, e.g., spell checkers
- Graphics manipulation
- Calligraphy
- Font design

Short Term Goals
- Find a vocabulary
- Find categories
- Classify systems
- Find common ground for reasoning
- Discuss current and future system capabilities
- Built up some abstractions
- Present results independent of software artefacts
Tactics

Don’t describe and focus on the functionality of 7th sub-sub-menu of 3rd subsystem’s dialog —

Intention of the creator is what counts!

- Identify underlying models
- Aim of classification demands decision criteria
- Find discriminators

Real Danger: How many angels can be on a 1sp rule?

Examples: Formatting Objects

“Lines” – New term needed
Abstraction: Sequence of characters along one direction

“Paragaphs” – There’s more to a term
E.g.: \TeX\ with its simple paragraph abstraction cannot handle
- Medieval scripts (inline paragraphs separated by \p)
- First fit (word-processor style paragraphs)

“Grids / Glue” – Intention of creator
- Systems can often be coerced to do both
- Creator had a clear intent—and we want to capture that
- Examples: QuarkXPress, \TeX\
Examples: Micro-Typography, Observers

Micro-typography: separation of concerns
- Punctuation
- Text input methods (alphabetic, syllabic, sinographemes)
- Segmentation (words, sub-words, . . . )
- Oft’ forgotten: text symbols

Observers: missing abstractions
- Observers in current systems are not systematic, most are very limited — e.g., page number of current text position, but no line number
- Missing: object model, with an observer interface

Categories for document representations
Interfaces to the outside:
- User interface style
- Character / glyph representation
- User views of the document
- Representation of the formatted document
- Import of non-native material

Example: Problems with Graphics Integration

Primary Site
- Application Data
- Files
- Database
- Configuration
- Application Programs
- System Programs
- Hardware

Backup Site
- Application Data
- Files
- Database
- Configuration
- Application Programs
- System Programs
- Hardware

Redundant Network Infrastructure
A Taxonomy of Automated Typesetting Systems
Joachim Schrod, Christopher A. Rowley, Christine Detig

Example: Text Flow and Graphics

The Brain: Style Sheets

Design interface
- Vocabulary: what is a style sheet
- Integration: relation of style sheets to documents
- Examples: Word — LaTeX — XSLT/FOP

What have we learnt?
- Far bigger than Joachim thought!
- Far more interesting than Chris thought!
- Far less current literature than Christine thought!

Wrap Up
- Classification is possible and sensible
- Classification has many different facets
- More than one way to format a document
- More than one way to model the formatting of a document
- New names are needed for old ideas
- New ideas are needed for old names
- Clarification of vocabulary is important
- Much work needed on classification and formalisation
- Much work needed on investigating current systems