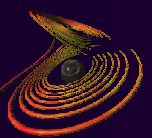
Digital Libraries: What Type of Documents Should We Be Dealing With??

D. Fellner

Braunschweig University of Technology, Germany

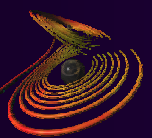
d.fellner@tu-bs.de



How much Information?

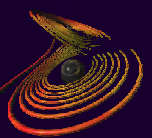
According to Lyman et al (Berkeley 2000):

- Annual production worldwide: *10^{18} bytes*
- Hardcopy only: *0,003%*
- Majority of information consisting of *Images, Animations, Sound, 3D Models* and *other numerical data*
- The (continuously increasing) major part of produced material is generated, archived, and exchanged in *digital form* – currently approx. *90% of total volume*.



A Digital Library ...

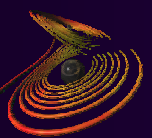
- ... provides text, images, animations, audio- and video-material in electronic form
  ☞ *electronic library*
- ... offers numerous library services independent from a particular location
  ☞ *virtual library*
- ... provides *efficient access* to *integrated content* and *services* through a *uniform interface* thereby providing substantially *more functionality than the sum of its components*



A Digital Library ...

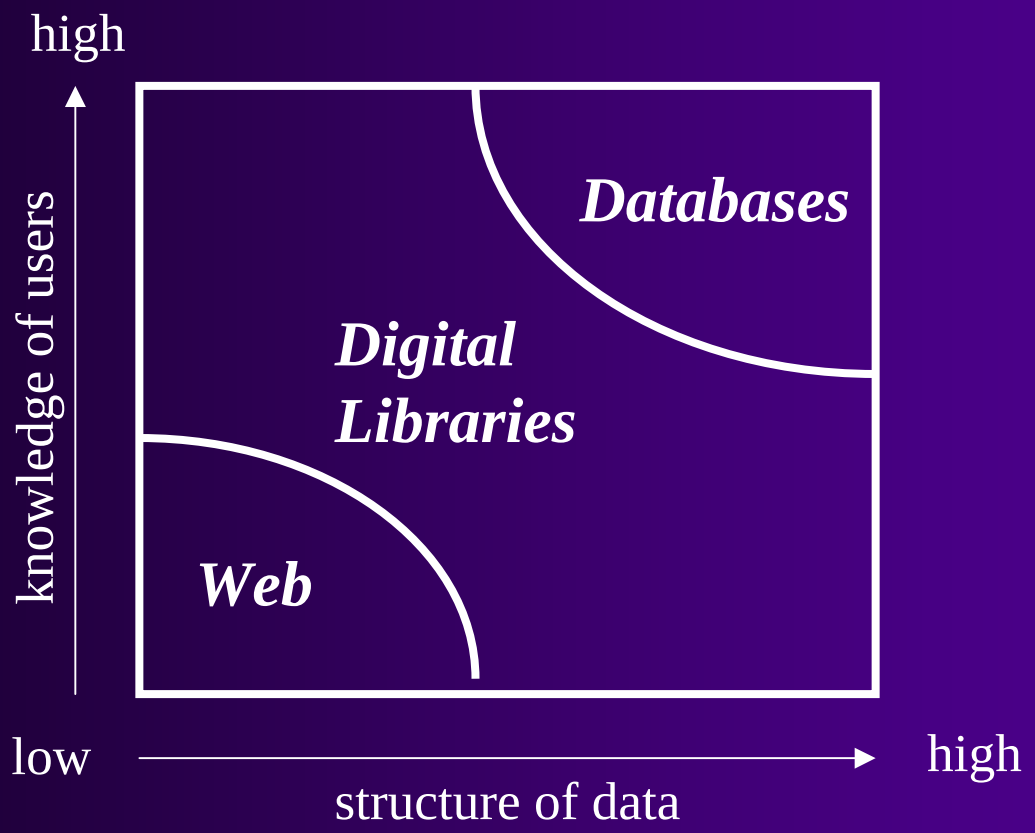
… addresses 2 related issues:

- *Digitizing all kind of media* results in an electronic library
- *Virtualizing services* results in a virtual library



Digital Library

[DELOS Meeting 2001:]

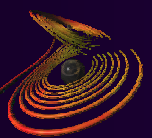


Digital Libraries ...

... according to Akscyn and Witten [DL‘99] will be one of the most important and most influential services/technologies of the 21. century

The reason being that DL‘s will enable/improve

- the *world-wide accessibility and semantic interlinking of mankind‘s knowledge*
- the *productivity and cooperation of knowledge workers*

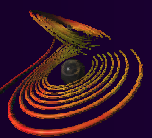


A Digital Library ...

... has to adopt a ‚generalized' view on the notion ‚*document*' being a collection of

- *Text*
- *Audio and Video*
- *Diagrams*, *Pictures*, *3D-Models*, *Animations*,
- *Raw Data* (e.g. from sensors toExcel-Sheets)
- or *dynamic e-learning material*

all in electronic form

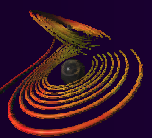


Strategic DL Research Initiative

DFG SPP 1041 – V³D²
Generalized Digital Documents

*acronym for: **V**erteilte **V**erarbeitung & **V**ermittlung **D**igitaler **D**okumente*
see http://graphics.tu-bs.de/V3D2

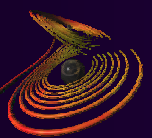
German Research Foundation
Starting 1998, 3 Phases
Funding volume/phase: ~4.5 Mio €
~ 40 Researches, ~ 22 Groups



DL's ⇔ Computer Graphics

Mutual benefits and challenges:

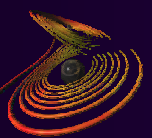
- Content Classification (of non-text documents)
- Retrieval and Search by Similarity
- Summarization
- Navigation (e.g. in distrib. 3D docs)
- Dissemination
- Linking (in non-text documents)
- …



Open Problems

- Content classification, Information extraction & indexing of non-text doc‘s:
  - content classification/analysis
  - search by similarity
  - type-based search
  - abstracting (in a general sense)

  for Audio, Video (+ shot-analysis & identification), Images, 3D Graphics (form-based search ~ digital mock-ups)

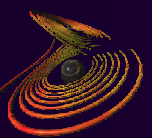


Open Problems

- Summarization of non-text doc‘s:

  Can be seen as the *overall term* to compaction and filtering.

  Comprises compaction and filtering aspects as well as data fusion (combining data of different type from potentially different sources) and should provide operations on the result set.



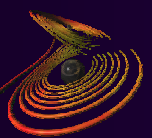
3D Documents

Goal:

complete semantic 3D-Model instead of projections in lower dimensions (image, section, animation, text) or structure-less collections of polygons

or

Preservation, Exploitation, but also Protection of semantic Information



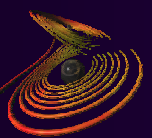
Representation of 3D Documents

Traditional Approach:

(hierarchical) polygonal meshes

Problems:

- *Loss of Structure*
- *Content-based handling (almost) impossible*
- *Inappropriate complexity measure*
  (sphere: (center,radius) $\leftrightarrow$ ∞ number of triangles)
- *Data compression very hard*
- *No object-specific LOD*
- *No protection against Reverse Engineering*



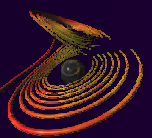
Representation of 3D Documents

New approach:

Progressive Combined BReps

Features:

- *Multi-resolution Data Structure*
- *Combines polygonal & freeform elements*
- *Control mesh can be modified dynamically by Euler-Operators*
- *Dynamic Tessellation at interactive rates*
- *Semantic LOD via ‚Euler Macros‘*



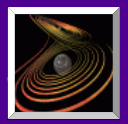
Progressive Combined BReps

Exploit semantic information:

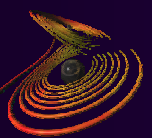
- *High-Level Culling and*
- *Dynamic adaptive Tessellation*

provide new level of quality and performance for visualization

E.g.: BioBrowser

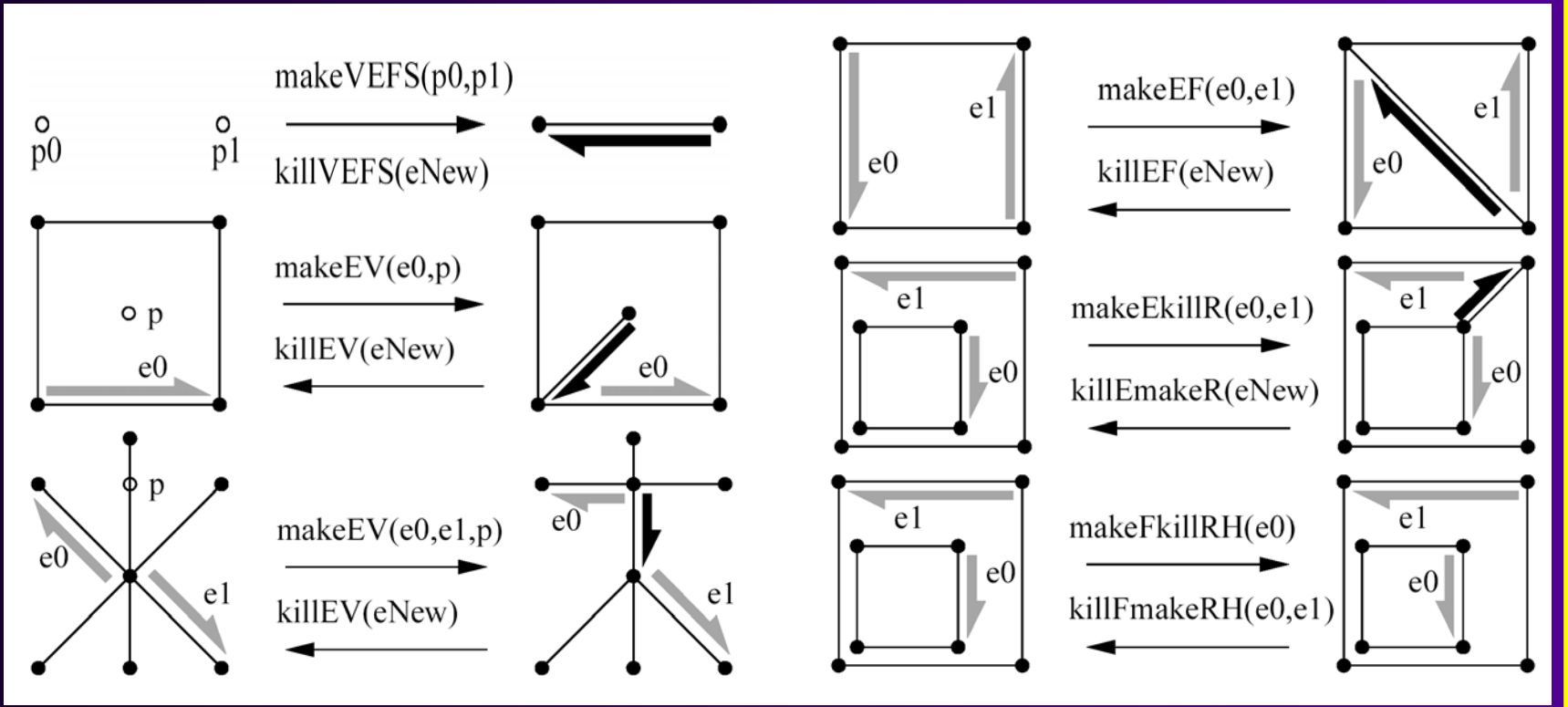


viewer for molecules in PDB-Format,
see http://graphics.tu-bs.de



Euler Operators

Invertible operators to manipulate meshes (insert/delete vertex/half-edges/face)



GML – Generative Modeling Language

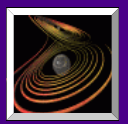
Encode & use semantic information:

Basic idea:

Separate Markup and Presentation

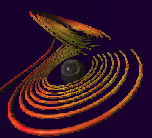
Similar to DTP

Example: Arcades



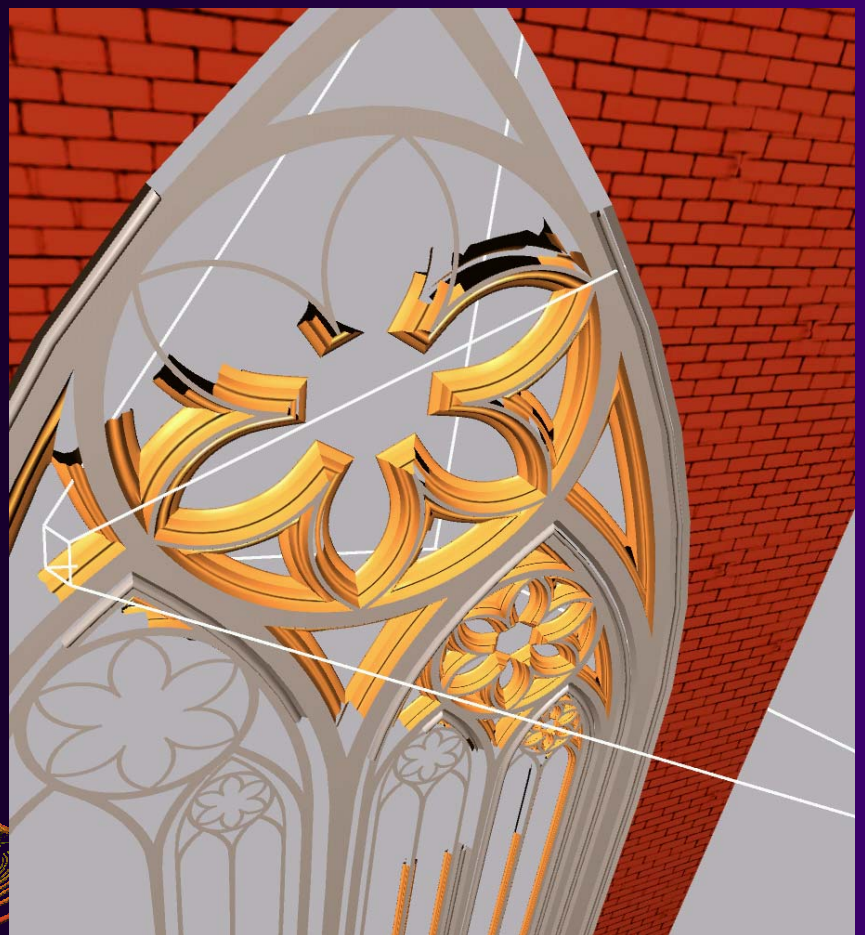
Euler Macros

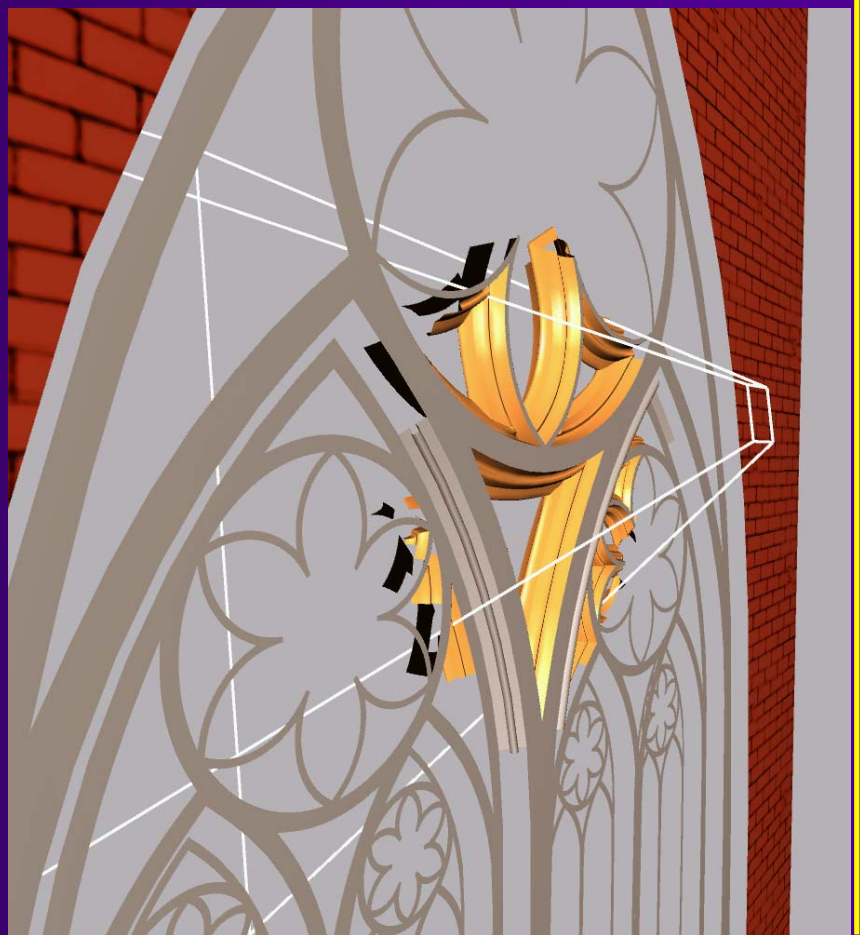
Browser Plugin



GML – Generative Modeling Language

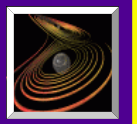
Example (contd): Gothic Style – a classical example for the use of parameterized models



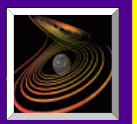


GML – Generative Modeling Language

E.g.: Cathedral in Cologne – *the* ultimate challenge



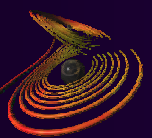
Pisa



Köln

Message to take home

- DL's contain only a fraction of the digital doc's they should
- Text is ok, but constitutes only a tiny fraction of the 1-2 exabyte (10^{18} byte) of relevant information [Lyman*00]
- Talk to the engineer next door:
  … *forget the text – the model is the thing that really counts* …



Thank You!

DFG Digital Library Research Initiative:

graphics.tu-bs.de/V3D2

Generative Modeling Project:

www.generative-modeling.org