



Application of UriGraph to Uniform Resource Identifier Design

Hrvoje Šimić

University of Zagreb

Faculty of Electrical Engineering and Computing

Department of Telecommunications



Overview

- Web **resources** and **URIs**
- **designing** URIs
 - use cases
 - requirements
 - recommendations
- **UriGraph**
 - resource identifier model
 - examples from the case study



Web Resource

- **source** of data
- **useful** information
- **available** through computer network
- **persistent**

Uniform Resource Identifier

- general term including (and replacing) **URLs**
- **exposed** to user
 - part of the Web user interface
- primary interest in **http-URI**
 - for identifying Web resources

http-URI anatomy

`http://example.org/a/b?x=1&y`

- **scheme**: http
- **host name** (and port)
- **path**
 - segments separated with slashes
- **query**
 - segments separated with ampersands
 - key and value separated with equals sign

A red decorative graphic consisting of a vertical line and two curved lines that intersect to form a stylized shape, possibly resembling a stylized 'U' or a bracket. It is positioned to the left of the title.

Use Cases for URIs

- hyperlink-enabled media
 - activating
- character-manipulative media
 - copy & paste
- other
 - reading
 - writing
 - memorizing
 - guessing
 - constructing

A red decorative graphic consisting of a vertical line and two curved lines forming a partial circle on the left side, positioned to the left of the title.

URI Design Requirements

- meaningfulness
- persistence
- good structure
- shortness
- readability
- memorizability
- pronounceability

Recommendations (1)

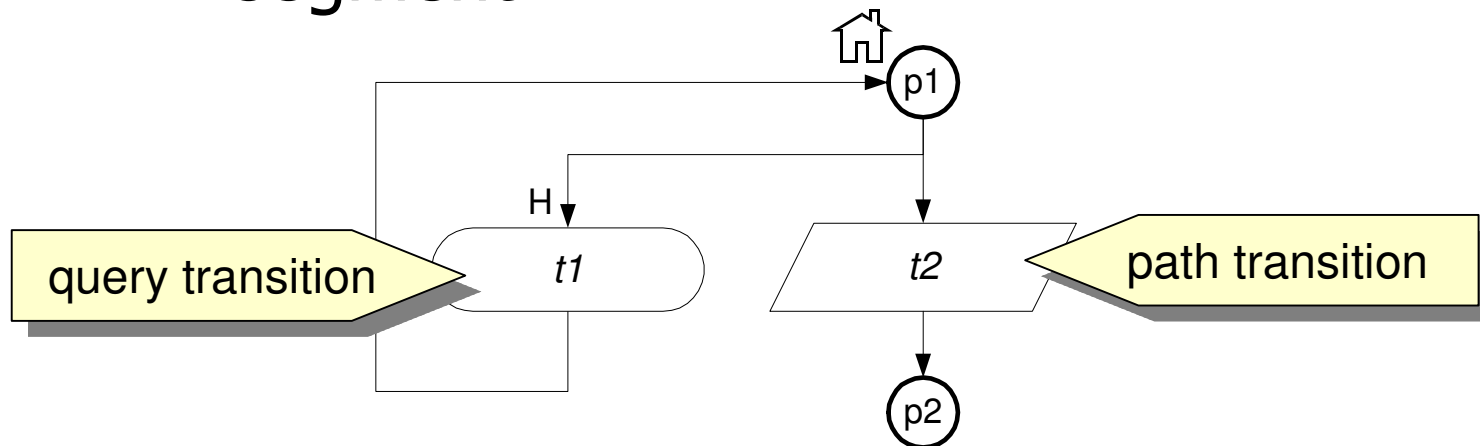
- some **lexical** limitations on path
 - letters and digits
 - hyphen, dot, slash
 - no underscore or comma
- **case insensitivity**
 - lowercase preferred
- **host names**
 - no “www” prefix
 - separating words with hyphens

Recommendations (2)

- **content** mistakes
 - irrelevant information
 - technical details
 - organizational information
 - changeable properties
 - redundant information
 - lack of information
- **multilingualism**
 - translated URIs for each language

UriGraph

- Web site structure model
- resource identifier model
- digraph with two types of nodes:
 - **places** represent classes of resources
 - **transitions** parse one path or query segment



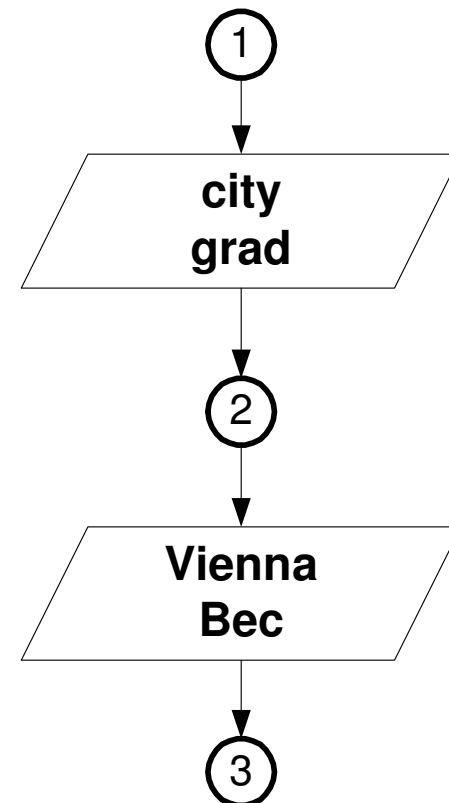


Pass

- transitions trim path or query segments upon traversal
- a **pass** tests a segment (or a part of a segment)
 - may be composite
- passes are implemented in Java
 - simplest: **fixed** pass
 - **variable** pass: tests segment against site content

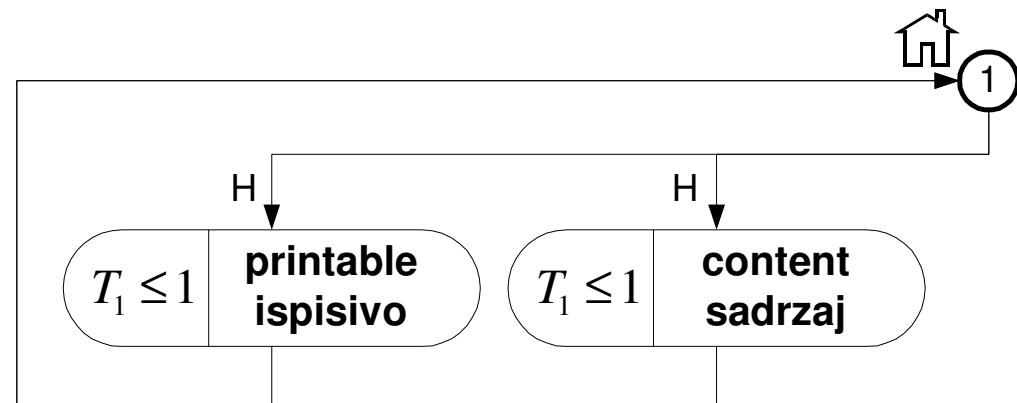
Multilingual pass

- a pass containing a list of subpasses for each language
- allowed:
 - grad
 - grad/Bec
 - city/Vienna
- but not:
 - grad/Vienna



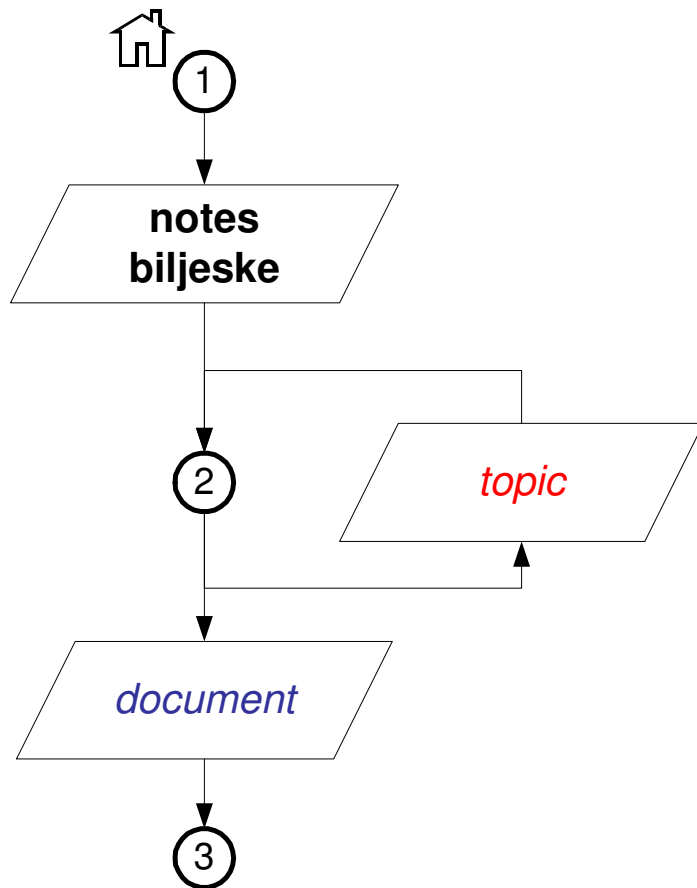
Traverse limit passes

- counting and limiting number of traversals through each transition



- allowed:
 - /
 - /?printable
- but not:
 - /?printable&content
 - /?printable&printable

Semi structured site



- a tree of topics
 - /notes
 - /notes/web/resource/general
 - /notes/web/020912-intro
- a loop enables a sequence of segments
- semi structured site
 - site structure depends largely on content



Conclusions

- URIs are important part of Web UI
- resource identifiers are foundations of site structure
- designing Web sites should start with URI design
- UriGraph can be efficiently used for URI design