SiSU - SiSU information Structuring Universe / Structured information, Serialized Units

FAQ - Frequently Asked/Answered Questions

Ralph Amissah

copy @ www.jus.uio.no/sisu/ [*]

Copyright © Ralph Amissah 2007, part of SiSU documentation, License GPL 3

Generated by SiSU [SiSU 0.59.0 of 2007w38/0] www.jus.uio.no/sisu

Copyright © 1997, current 2007 Ralph Amissah, All Rights Reserved.

SiSU is software for document structuring, publishing and search (with object citation numbering), www.sisudoc.org

SiSU is released under GPL 3 or later, <http://www.fsf.org/licenses/gpl.html>.

Document information:
sourcefile sisu_faq.sst
Generated by SiSU www.jus.uio.no/sisu
version information: SiSU 0.59.0 of 2007w38/0

For alternative output formats of this document check:

<http://www.jus.uio.no/sisu/sisu_faq/sisu_manifest.html>

Contents

SiSU - SiSU information Structuring Universe / Structured information, Serialized Units - FAQ - Frequently Asked/Answered Questions,
Ralph Amissah [1]

1. FAQ - Frequently Asked/Answered Questions <small>2</small>

1.1 Why are urls produced with the -v (and -u) flag that point to a web server on port <small>3</small> 8081?

Try the following rune: <small>4</small>

- sisu -W <small>5</small>

This should start the ruby webserver. It should be done after having produced some output as it <small>6</small> scans the output directory for what to serve.

1.2 I cannot find my output, where is it? <small>7</small>

The following should provide help on output paths: <small>8</small>

- sisu –help env <small>9</small>
- sisu -V [same as the previous command] <small>10</small>
- sisu –help directory <small>11</small>
- sisu –help path <small>12</small>
- sisu -U [filename] <small>13</small>
- man sisu <small>14</small>

1.3 I do not get any pdf output, why? <small>15</small>

SiSU produces LaTeX and pdflatex is run against that to generate pdf files. <small>16</small>

If you use **Debian** the following will install the required dependencies <small>17</small>

- aptitude install sisu-pdf <small>18</small>

the following packages are required: tetex-bin, tetex-extra, latex-ucs <small>19</small>

1.4 Where is the latex (or some other interim) output? <small>20</small>

Try adding -M (for maintenance) to your command flags, e.g.: <small>21</small>

- sisu -HpMv [filename] <small>22</small>

this should result in the interim processing output being retained, and information being pro- <small>23</small> vided on where to find it.

- sisu –help directory <small>24</small>
- sisu –help path <small>25</small>

should also provide some relevant information as to where it is placed. <small>26</small>

1.5 Why isn't SiSU markup XML

I worked with text and (though I find XML immensely valuable) disliked noise ... better to sidestep the question and say:

SiSU currently "understands" three XML input representations - or more accurately, converts from three forms of XML to native **SiSU** markup for processing. The three types correspond to SAX (structure described), DOM (structure embedded, whole document must be read before structure is correctly discernable) and node based (a tree) forms of XML document structure representation. Problem is I use them very seldom and check that all is as it should be with them seldom, so I would not be surprised if something breaks there, but as far as I know they are working. I will check and add an XML markup help page before the next release. There already is a bit of information in the man page under the title **SiSU** VERSION CONVERSION

sisu –to-sax [filename/wildcard]

sisu –to-dom [filename/wildcard]

sisu –to-node [filename/wildcard]

The XML should be well formed... must check, but lacks sensible headers. Suggestions welcome as to what to make of them. [For the present time I am satisfied that I can convert (both ways) between 3 forms of XML representation and **SiSU** markup].

sisu –from-xml2sst [filename/wildcard]

1.6 LaTeX claims to be a document preparation system for high-quality typesetting. Can the same be said about SiSU?

SiSU is not really about type-setting.

LaTeX is the ultimate computer instruction type-setting language for paper based publication.

LaTeX is able to control just about everything that happens on page and pixel, position letters kerning, space variation between characters, words, paragraphs etc. formula.

SiSU is not really about type-setting at all. It is about a lightweight markup instruction that provides enough information for an abstraction of the documents structure and objects, from which different forms of representation of the document can be generated.

SiSU with very little markup instruction is able to produce relatively high quality pdf by virtue of being able to generate usable default LaTeX; it produces "quality" html by generating the html directly; likewise it populates an SQL database in a useful way with the document in object sized chunks and its meta-data. But **SiSU** works on an abstraction of the document's structure and content and custom builds suitable uniform output. The html for browser viewing and pdf for paper viewing/publishing are rather different things with different needs for layout - as indeed is what is needed to store information in a database in searchable objects.

The pdfs or html produced for example by open office based on open document format and

other office/word processor suits usually attempt to have similar looking outputs - your document rendered in html looks much the same, or in pdf... sisu is less this way, it seeks to have a starting point with as little information about appearance as possible, and to come up with the best possible appearance for each output that can be derived based on this minimal information.

Where there are large document sets, it provides consistency in appearance in each output format for the documents. 42

The excuse for going this way is, it is a waste of time to think much about appearance when working on substantive content, it is the substantive content that is relevant, not the way it looks beyond the basic informational tags - and yet you want to be able to take advantage of as many useful different ways of representing documents as are available, and for various types of output to to be/look as good as it can for each medium/format in which it is presented, (with different mediums having different focuses) and **SiSU** tries to achieve this from minimal markup. 43

1.7 How do I create GIN or GiST index in Postgresql for use in SiSU 44

This at present needs to be done "manually" and it is probably necessary to alter the sample search form. The following is a helpful response from one of the contributors of GiN to Postgresql Oleg Bartunov 2006-12-06: 45

"I have tsearch2 slides which introduces tsearch2 http://www.sai.msu.su/~ megera/wiki/tsearch2slides 46

FTS in PostgreSQL is provided by tsearch2, which should works without any indices (GiST or GIN) ! Indices provide performance, not functionality. 47

In your example I'd do (simple way, just for demo): 48

0. compile, install tsearch2 and load tsearch2 into your database 49

cd contrib/tsearch2; makemakeinstallmake installcheck; psql DB < tsearch2.sql 50

1. Add column fts, which holds tsvector 51

alter table documents add column fts tsvector; 52

2. Fill fts column 53

update document set fts = to_tsvector(clean); 54

3. create index - just for performance ! 55

create index fts_gin_idx on document using gin(fts); 56

4. Run vacuum 57

vacuum analyze document; 58

That's all. 59

Now you can search: 60

select lid, metadata_tid, rank_cd(fts, q,2)as rank from document, plainto_tsquery('markup syn- tax') q where q @@ fts order by rank desc limit 10;   61

1.8 Where is version 1.0?   62

SiSU works pretty well as it is supposed to. Version 1.0 will have the current markup, and   63
directory structure. At this point it is largely a matter of choice as to when the name change is
made.

The feature set for html,[1] LaTeX/pdf and opendocument is in place. XML, and plaintext are in   64
order.

html and LaTeX/pdf may be regarded as reference copy outputs   65

With regard to the populating of sql databases (postgresql and sqlite), there is a bit to be   66
done.

We are still almost there.   67

---

[1] html w3c compliance has been largely met.

DOCUMENT INFORMATION (METADATA)

Metadata

Document Manifest @

〈http://www.jus.uio.no/sisu/sisu_manual/sisu_faq/sisu_manifest.html〉

Dublin Core (DC)

DC tags included with this document are provided here.

DC Title: SiSU - SiSU information Structuring Universe / Structured information, Serialized Units - FAQ - Frequently Asked/Answered Questions

DC Creator: Ralph Amissah

DC Rights: Copyright (C) Ralph Amissah 2007, part of SiSU documentation, License GPL 3

DC Type: information

DC Date created: 2006-09-06

DC Date available: 2006-09-06

DC Date issued: 2006-09-06

DC Date modified: 2007-09-16

DC Date: 2007-09-16

Version Information

Sourcefile: sisu_faq.sst

Filetype: SiSU text 0.57

Sourcefile Digest, MD5(sisu_faq.sst)= 4ba78b1270b9d994054c1ff06d58765d

Skin_Digest: MD5(/home/ralph/grotto/theatre/dbld/sisu-dev/sisu/data/doc/sisu/sisu_markup_samples/sisu_mar 20fc43cf3eb6590bc3399a1aef65c5a9

Generated

Document (metaverse) last generated: Sun Sep 23 04:11:02 +0100 2007

Generated by: SiSU 0.59.0 of 2007w38/0 (2007-09-23)

Ruby version: ruby 1.8.6 (2007-06-07 patchlevel 36) [i486-linux]

Information on this document copy and an unofficial List of Some web related information and sources

"Support Open Standards and Software Libre for the Information Technology Infrastructure" RA

Information on this document copy www.jus.uio.no/sisu/

Generated by SiSU found at www.jus.uio.no/sisu [SiSU 0.59.0 2007w38/0] www.sisudoc.org. SiSU is software for document structuring, publishing and search (using SiSU: object citation numbering, markup, meta-markup, and system) Copyright © 1997, current 2007 Ralph Amissah, All Rights Reserved.

SiSU is released under GPL 3 or later (www.fsf.org/licenses/gpl.html).

W3 since October 3 1993 *SiSU* SiSU 1997, current 2007.

SiSU presentations at **www.jus.uio.no/sisu/**

SiSU - SiSU information Structuring Universe / Structured information, Serialized Units **pdf** versions can be found at:

http://www.jus.uio.no/sisu/sisu_faq/portrait.pdf

http://www.jus.uio.no/sisu/sisu_faq/landscape.pdf

SiSU - SiSU information Structuring Universe / Structured information, Serialized Units **html** versions may be found at:

http://www.jus.uio.no/sisu/sisu_faq/toc.html or

http://www.jus.uio.no/sisu/sisu_faq/doc.html

SiSU Manifest of document output and metadata may be found at:

http://www.jus.uio.no/sisu/sisu_faq/sisu_manifest.html

SiSU found at: **www.jus.uio.no/sisu/**

Links that may be of interest at SiSU and elsewhere:

SiSU Manual

http://www.jus.uio.no/sisu/sisu_manual/

Book Samples and Markup Examples

http://www.jus.uio.no/sisu/SiSU/2.html

SiSU @ Wikipedia

http://en.wikipedia.org/wiki/SiSU

SiSU @ Freshmeat

http://freshmeat.net/projects/sisu/

SiSU @ Ruby Application Archive

http://raa.ruby-lang.org/project/sisu/

SiSU @ Debian

http://packages.qa.debian.org/s/sisu.html

SiSU Download

http://www.jus.uio.no/sisu/SiSU/download.html

SiSU Changelog

http://www.jus.uio.no/sisu/SiSU/changelog.html

SiSU help

http://www.jus.uio.no/sisu/sisu_manual/sisu_help/

SiSU help sources

http://www.jus.uio.no/sisu/sisu_manual/sisu_help_sources/

SiSU home:
www.jus.uio.no/sisu/