Inline Tag Handling

Dani Rosàs
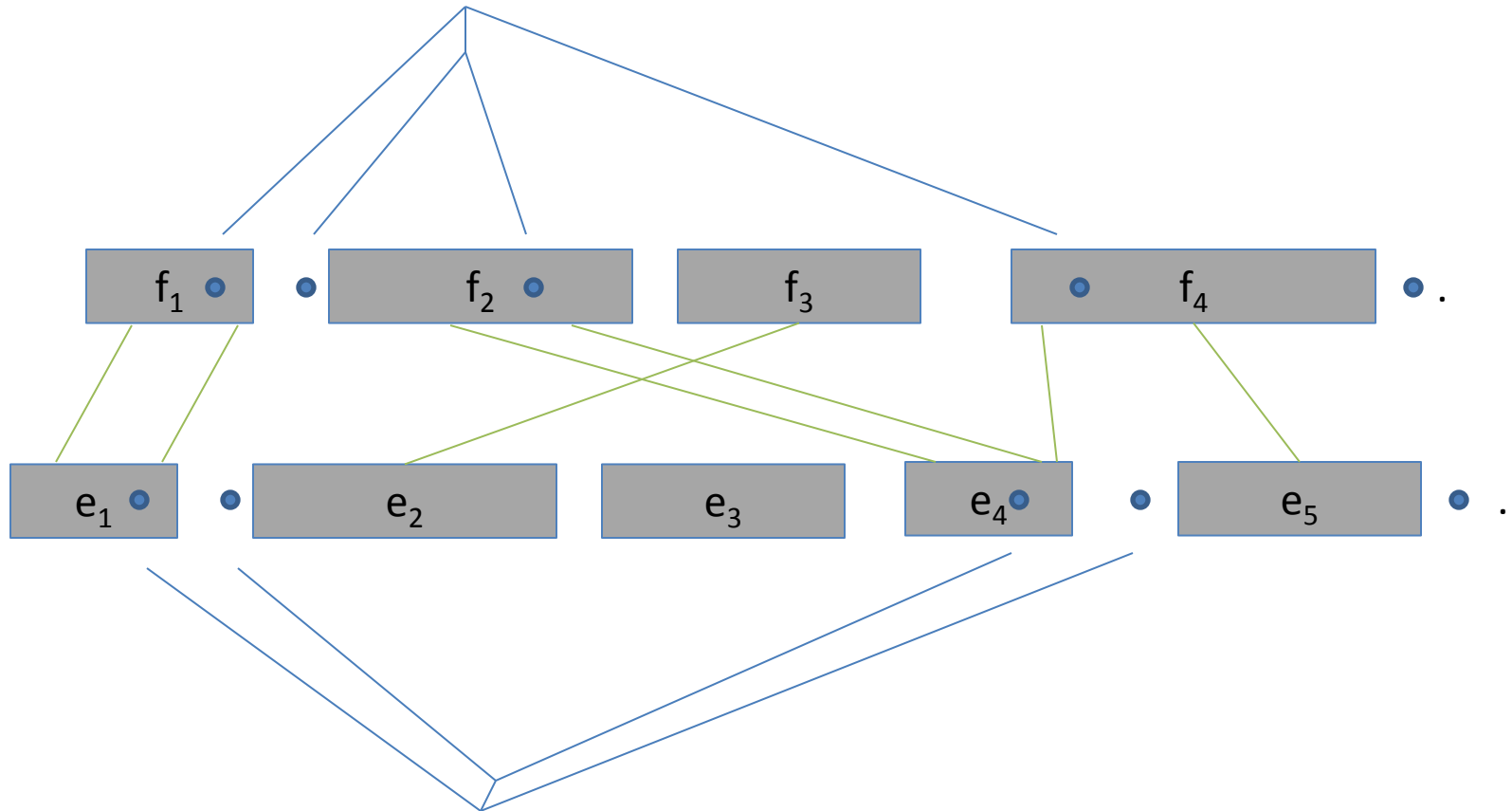
Jacek Skarbek

Tomas Fulajtar

Zdenka Závůrková

Achim Ruopp

Using Word Alignment Information to Place Inline Formatting Tags

Inline Tagging Done

- Coded tag insertion based on word alignment
- Looked at related solution developed by Christian Buck
- Test data for evaluation
  - Need test data with inline formatting tags
  - Aligned product documentation from IBM available on the web
  - Processing documentation from GNU project in DocBook format
- Built small En-Es Moses MT system with truecasing
  - From WMT 2013 development data
  - Doesn't need to be good, but some reordering needed
  - Couldn't use pre-built v1.0 pre-built En-Es phrase-based system
- Integrated Truecaser in overall process
  - Moses tools not flushing buffers!

Inline Tagging Further Done

- prepared working solution based on word alignment

- Evaluated different approaches :

Method	BLEU
Word aligned	17.14
Phrase aligned	16.71

- Manuall check: we saw 113 corrections done

Inline Tagging comparison

For configuration instructions, see <g id="link" name="bstgbqn">***Local Configuration Options***</g>.

- Phrase based alignment:

Para configuración instrucciones, consulte ***opciones de configuración*** <g id="link" name="bstgbqn"></g>***local***.

- Word based alignment:

Para configuración instrucciones, consulte <g id="link" name="bstgbqn"> ***opciones de configuración local***</g>.

Inline Tagging TBD

- Integrate solution based on word alignment
- Evaluate different approaches and potentially merge
  - Christian Buck's tags4moses solution (Matecat)
  - EC placeholder handling
  - Others?
- Get code into Moses github