

Computational Linguistics I, Winter 2006. Marcus Kracht

To be submitted: Friday, February 24, 2006.

- [A 5.1] A simple HTML-Tag is as follows: it is of the form `<h3 options>`, where *options* is one or more of the following: `fontname="fontname"`;;, where *fontname* is a sequence of letters and digits; `color="colorname"`;;, and a *colorname* is a sequence of the form `#xxxxxx`, where each *x* is a hexadecimal number: this means that it is either a digit or a letter from **a** to **f**. Colournames and tags are case insensitive. First, calculate a regular expression (in ordinary notation) that accepts all and only the well-formed tags.
- [A 5.2] Determine which OCaml string represents the regular expression of the previous exercise. Run it on a few examples.
- [A 5.3] Write a regular expression that eliminates the optional fields from the HTML-Tag (via the use of the `Str.global_replace` function or its kin).
- [A 5.4] As in the previous exercise, write a regular expression that allows to translate singular nouns of English into their plural form. As before, do not try to be exhaustive, but show some representative cases.