From Natural Tokens to Natural Trees in Source-Code

Bringing Structure to Naturalness: On the Naturalness of ASTs

PĂRȚACHI Profir-Petru, 杉山麿人
(国立情報学研究所)

Previous research have shown source-code to be statistically predictable as a sequence of tokens. Is the same still true for structured views of source-code such as Abstract Syntax Trees?

Hindle et al.[2] showed that source-code at the token level is more predictable than standard English. It is also intuitive to assume that adding information can only simplify the prediction task, thus recent work employ tree or graph views of source-code. However, the hypothesis that such views should be “natural”, i.e. statistically predictable, has not been tested.

Hindle et al.[2] showed that source-code at the token level is more predictable than standard English. It is also intuitive to assume that adding information can only simplify the prediction task, thus recent work employ tree or graph views of source-code. However, the hypothesis that such views should be “natural”, i.e. statistically predictable, has not been tested.

Hindle et al.[2] showed that source-code at the token level is more predictable than standard English. It is also intuitive to assume that adding information can only simplify the prediction task, thus recent work employ tree or graph views of source-code. However, the hypothesis that such views should be “natural”, i.e. statistically predictable, has not been tested.

Hindle et al.[2] showed that source-code at the token level is more predictable than standard English. It is also intuitive to assume that adding information can only simplify the prediction task, thus recent work employ tree or graph views of source-code. However, the hypothesis that such views should be “natural”, i.e. statistically predictable, has not been tested.

Hindle et al.[2] showed that source-code at the token level is more predictable than standard English. It is also intuitive to assume that adding information can only simplify the prediction task, thus recent work employ tree or graph views of source-code. However, the hypothesis that such views should be “natural”, i.e. statistically predictable, has not been tested.

Hindle et al.[2] showed that source-code at the token level is more predictable than standard English. It is also intuitive to assume that adding information can only simplify the prediction task, thus recent work employ tree or graph views of source-code. However, the hypothesis that such views should be “natural”, i.e. statistically predictable, has not been tested.

Hindle et al.[2] showed that source-code at the token level is more predictable than standard English. It is also intuitive to assume that adding information can only simplify the prediction task, thus recent work employ tree or graph views of source-code. However, the hypothesis that such views should be “natural”, i.e. statistically predictable, has not been tested.

Hindle et al.[2] showed that source-code at the token level is more predictable than standard English. It is also intuitive to assume that adding information can only simplify the prediction task, thus recent work employ tree or graph views of source-code. However, the hypothesis that such views should be “natural”, i.e. statistically predictable, has not been tested.

Hindle et al.[2] showed that source-code at the token level is more predictable than standard English. It is also intuitive to assume that adding information can only simplify the prediction task, thus recent work employ tree or graph views of source-code. However, the hypothesis that such views should be “natural”, i.e. statistically predictable, has not been tested.

Hindle et al.[2] showed that source-code at the token level is more predictable than standard English. It is also intuitive to assume that adding information can only simplify the prediction task, thus recent work employ tree or graph views of source-code. However, the hypothesis that such views should be “natural”, i.e. statistically predictable, has not been tested.

Hindle et al.[2] showed that source-code at the token level is more predictable than standard English. It is also intuitive to assume that adding information can only simplify the prediction task, thus recent work employ tree or graph views of source-code. However, the hypothesis that such views should be “natural”, i.e. statistically predictable, has not been tested.

Hindle et al.[2] showed that source-code at the token level is more predictable than standard English. It is also intuitive to assume that adding information can only simplify the prediction task, thus recent work employ tree or graph views of source-code. However, the hypothesis that such views should be “natural”, i.e. statistically predictable, has not been tested.