



MEDFORD Language Server

Liam Strand and Lenore J. Cowen
Department of Computer Science
Tufts University, Medford, MA

Tufts
UNIVERSITY

Department of
Computer Science

MOTIVATION

In every field, scientists must collect some type of data. These data can be as varied as material strength measurements to software performance statistics. To be useful in a research setting, these data must be contextualized. It is important, for example, to know exactly what hardware was used to take a photograph, or what versions of software packages were used to perform data analysis. We call this “data about data” metadata, and it is critical to allow other researchers to validate results in the peer-review and post-publication feedback process.

This information is often recorded in handwritten notes, which is an extremely flexible and intuitive format. But, those notes must be converted into a digital format for sharing and database submissions, so that computers can parse the metadata. These digital formats are highly structured, but they are extremely challenging to write for a non-specialist.

The MEDFORD metadata markup language combines the best of both of these options. The format is as flexible as handwritten notes, yet as machine-parsable as a highly-structured RDF file. Furthermore, a MEDFORD file can be converted into other formats for database submissions. [1]

Though this new language is a huge step forward, there is a learning curve to quickly writing in MEDFORD. This new work aims to allow users to seamlessly write in MEDFORD, receiving instant feedback on their correctness, and discovering new language features.



REFERENCES

[1] Shpilker, Polina, John Freeman, Hailey McKelvie, Jill Ashey, Jay-Miguel Fonticella, Hollie Putnam, Jane Greenberg, Lenore J. Cowen, Alva Couch, and Noah M. Daniels, *MEDFORD: A human and machine readable metadata markup language*, <https://doi.org/10.48550/ARXIV.2204.09610>, 2022

FEATURES

ERROR CHECKING

```
field required MEDFORD(missing_field)
View Problem No quick fixes available
@Software_Ref MEDFORD Parser Source Code
@Software_Ref-Filename medford
@Software_Ref-URI https://github.com/TuftsBCB/MEDFORD
```

```
invalid date format MEDFORD(value_error)
View Problem No quick fixes available
@Date 2022/08/02
@Date-Note First medford-vscode release!
```

SYNTAX HIGHLIGHTING

```
# Metadata about the software we've written
@Software_Primary mfdls
@Software_Primary-Path ./mfdls
@Software_Primary-Version 0.1.0
@Software_Primary-Type Language Server

# Metadata about the contributors
`@tufts_cs Department of Computer Science, Tufts University,
177 College Ave, Medford MA, 02155
`@uri_cs_stat Department of Computer Science and Statistics,
University of Rhode Island, Kingston, RI 02881

@Contributor Liam Strand
@Contributor-Association `@tufts_cs
@Contributor-Email Liam.Strand@tufts.edu
@Contributor-Role Project Lead
```

IMPLEMENTATION

Text Editor



“Please autocomplete here!”
“The document changed!”
“Here are some errors...”
“Here are the completions...”

Language Server



“What tokens are defined?”
“Please parse this text...”
“I found these errors...”
“We are using these tokens...”

MEDFORD Parser



CHALLENGES

One of the primary challenges was that the MEDFORD parser outputs custom error objects when it detects syntactic or semantic errors. These objects needed to be converted into Language Server Protocol diagnostic objects in order to be sent to the text editor.

Another challenge was that the text editor and language server communicate using standard input and output. This means that any accidental output pollutes this communication channel and breaks the entire setup. Many steps were taken throughout the development process to prevent these erroneous outputs.

FUTURE WORK

Though the language server is functionally complete, it must continue to evolve with the MEDFORD language. We are currently discussing and testing supporting a “linking” token that allows sections of metadata to reference other sections.

Other future work includes the preparation and submission of a tutorial paper that introduces new users to the MEDFORD language and explains how the language server can make the onboarding process more seamless.

Finally, we hope to introduce a broad spectrum of researchers to the Language, help them use the support software, and get feedback and suggestions.

ACKNOWLEDGEMENTS

Thank you to all of my co-authors and collaborators:
Context-Sensitive Editing for the MEDFORD Metadata Language, Liam Strand(Tufts University), Andrew Powers(University of Rhode Island), Polina Shpilker(Tufts University), Alva Couch(Tufts University), Lenore J. Cowen(Tufts University), and Noah M. Daniels(University of Rhode Island), Manuscript in Preparation.

My collaborators and I would like to extend our gratitude to Cece Crumlish(Tufts University) for creating our beautiful logo (visible in the left column).

I would like to thank the Fowler Family Endowed Summer Scholars Fund for supporting this research.