Comparative Javascript Toolkit Review in Support of Students for CS 410 FA21

Intro

This information is current as of November 2021. For our course project, my group chose to implement an Intelligent Browsing extension to assist fellow students in their employment search (#1). Ideally, the extension could run without the need of maintaining server resources, therefore we chose to research options to run entirely in the browser. As a necessary first step in that direction, we needed to know which libraries are available in pure Javascript. Compiled implementations claiming faster throughput (for server side execution) were discovered in the process, but are not the primary intention of this review.

The review takes inspiration from guidelines specified by the Journal of Open Software (#2, #3). Several JS ML frameworks were listed on these blog pages (#4, #5). Others were found through referral, general google searching and references in YT video tutorials.

Review Criteria

- **Year of Last Update**
- **Software license** OSI approved license in repository from https://opensource.org/licenses
- **Documentation** Sufficient documentation to understand functionality.
- **A statement of need** Authors clearly state the target audience and the problem(s) being solved.
- **Installation instructions** Clear dependencies and automated package installation.
  - Good: gemfile, npm, package.json or equivalent
  - OK: A list of dependencies to install
  - Bad (not acceptable): Reliance on other software not listed by the authors
- **Example usage** examples included to solve real-world problems
- **API documentation** Subjective evaluation of API documentation
- **Tests** Automated test suite or core functionality is strongly encouraged.
  - Good: An automated test suite hooked up to an external service such as Travis-CI or similar
  - OK: Documented manual steps that can be followed to check the expected functionality of the software (e.g. a sample input file to assert behavior)
  - Bad (not acceptable): No way for you the reviewer to check whether the software works
- **Community guidelines** There should be clear guidelines for third-parties wishing to: Contribute to the software, report issues or problems with the software, or seek support
- **Server Run Option**
- **Browser Run Option**
- **Language Support** English or Multiple
- **Features | Functionality** Topic list of features or utility functions available in the package. Not comprehensive.

Packages Reviewed

Name	GitHub	WWW	Description
NLP.js	https://github.com/axa-group/nlp.js/		An NLP library for building bots, with entity extraction, sentiment analysis, automatic language identification and more.
Compromise	https://github.com/spencermountain/compromise		Modest natural-language processing, tiny size.
Neuro	https://github.com/intelligo-mn/neuro	https://neuro.js.org/	ML Library for building AI assistants and chat-bots. Work in Progress.
Natural	https://github.com/NaturalNode/natural	http://naturalnode.github.io/natural/	Broad range of functions for national language processing.
SpaCy-NLP	https://github.com/kengz/spacy-nlp		Expose SpaCy NLP text parsing to Nodejs via SocketIO
HuggingFace	https://github.com/huggingface/tokenizers		Fast State-of-the-Art Tokenizers optimized for Research and Production
Sentiment	https://github.com/thisandagain/sentiment		AFINN-based sentiment analysis
Wink-NLP-utils	https://github.com/winkjs	http://winkjs.org	Developer friendly NLP
lda	https://github.com/primaryobjects/lda		LDA Topic Modeling for Nodejs
FastText	https://github.com/facebookresearch/fastText/ https://github.com/loretoparisi/fasttext.js/	https://fasttext.cc/	Library for fast text representation and classification. FastText for Nodejs
Gensim	https://github.com/RaRe-Technologies/gensim	https://radimrehurek.com/gensim/index.html	Topic modeling for humans
GENRE	https://github.com/facebookresearch/GENRE		Autoregressive Entity Retrieval

Features + Functionality

Name	Features + Functionality
NLP.js	QnA, NER, NN, emoji, similarity, NLU, React Native, Language Guessing, Similar Search, Duckling - Entity Extraction, Native Entity Extraction, Sentiment, Slot Filling, Bot Framework.
Compromise	Slot Match, Verb Tense, Noun Plurality, Plain-text Numbers, Topic Query, Contractions, POS, Syllables
Neuro	NN Binary Classification, Feature Engineering, Multi-label classification, Cross-Validation, SVM Wrapper
Natural	Tokenizer, Distance (hamming, jaro, dice), Stemmer, Bayes Classifier, LR Classifier, Max Entropy, Sentiment, Inflectors, Phonetics, N-grams, TF-IDF, Tries, Edge Weighted Graph, Shortest Path Tree, Longest Path Tree, Wordnet, Spellcheck, POS
SpaCy-NLP	Subset of SpaCy (POS, Syntax, NER)
HuggingFace	Extensive, but limited to wrappers for Javascript
Sentiment	Sentiment Text, Sentiment - Emoji
Wink-NLP-utils	36 functions including: BM25, NER, Negation, Sentence Boundary, Chatbot, POS, Tokenizer, Pattern Match, Sentiment, Visualization, Similarity, Perceptron, Regression Tree etc
lda	LDA Topic
FastText	Subset of Python FastText
Gensim	Topic, Word2Vec, Embeddings, LDA, LSA, Random Projections, HDP
GENRE	Autoregressive Entity Retrieval

Review Results

Name	Software License	Last Updated	Docs	Statement of Need	Installation Instructions	Example Usage	API Docs	Contributors	Used by (per GitHub)	Community Guidelines	Server Package	Browser Package	Language Support	Tests
NLP.js	MIT	2021	YES	YES	GOOD	YES	YES	59	-	YES	YES	YES	multi	OK+
Compromise	MIT	2021	YES	YES	GOOD	YES	YES	87	1200+	YES	YES	YES	English	OK+
Neuro	LGPL	2021	YES	YES	GOOD	YES	YES	7	-	YES	YES	NO	English	OK+
Natural	MIT	2021	YES	YES	GOOD	YES	YES	126	10000+	YES	YES	YES	multi	OK+
SpaCy-NLP	MIT	2020	YES	YES	OK	YES	YES	8	43	NO	YES	NO	multi	OK-
HuggingFace	Apache 2.0	2021	YES	YES	GOOD	YES	YES	48	7300+	YES	YES	NO	multi	GOOD
Sentiment	MIT	2019	YES	YES	GOOD	YES	YES	18	4000+	YES	YES	YES	English	GOOD
Wink-NLP-utils	MIT	2021	YES	YES	GOOD	YES	YES	5	40	YES	YES	YES	English	OK+
Ida	Apache 2.0	2017	YES	YES	GOOD	YES	YES	5	162	NO	YES	YES	English	OK-
FastText	MIT MIT	2020	YES	YES	OK	YES	YES	49 3	2200+ -	YES NO	YES	NO	multi	GOOD
Gensim	LGPL	2021	YES	YES	OK	YES	YES	390	27400+	YES	YES	NO	English	GOOD
GENRE	CC BY-NC 4.0	2021	YES	YES	OK	YES	YES	3	-	YES	YES	NO	multi	OK+

Conclusion

Each student/group can draw their own conclusions as to which package might be useful to them. A few columns worth noting are Last Updated, Installation, Contributed+UsedBy, Browser Package and Tests.

A few packages have not been updated this year, which could suggest that support is waning. If Installation instructions are GOOD, then it should require less effort for a client side Javascript setup compared to a package install of OK. The Contributed+UsedBy numbers had a staggering discrepancy in some cases, e.g. Gensim by over 27k and Natural by over 10k users. However, Gensim does not provide browser side JS support. About half the packages reviewed support browser side implementation, which is pretty strong if a group wishes to implement a browser extension that doesn't require setup of a server side API. Only a handful of packages claimed full Continuous Integration testing for each release. However, a testing harness was found for the majority of packages.

Please email me (rahuls11@illinois.edu) or message me on Slack (@RahulSinghalChicago) if you find errors in this review, have suggestions for other packages to include or any other comments.

References
1. https://github.com/RahulSinghalChicago/CourseProject/blob/main/Project%20Proposal.pdf
2. https://joss.readthedocs.io/en/latest/review_criteria.html
3. https://joss.readthedocs.io/en/latest/review_checklist.html
4. https://machinelearningknowledge.ai/javascript-libraries-for-machine-learning-deep-learning-nlp-and-computer-vision-you-did-not-know-exist/
5. https://openbase.com/categories/js/best-nodejs-nlp-libraries