

TM-Town open sources sentence segmentation library

TM-Town has open sourced a rule-based sentence boundary detection library written in Ruby.

Lewes, DE, January 8th, 2015 - TM-Town has open sourced its sentence segmentation library. The library, called [Pragmatic Segmenter](#), is a rule-based sentence boundary detection library written in Ruby that works out-of-the-box across many languages. Pragmatic Segmenter is very accurate and beats out other popular sentence segmentation tools in edge case tests.

Kevin Dias, TM-Town's developer explains, "As TM-Town benefits from a lot of open-source technology I have decided to open source TM-Town's sentence segmentation library in hopes that it might benefit the community."

On TM-Town translators upload documents across many different fields of expertise and many languages. As such, it is important that TM-Town's segmentation engine be able to handle many different languages as well as deal with potentially ill-formatted content (i.e. text imported from PDFs often has line-breaks that fall in the middle of sentences).

"TM-Town needed a tool that could work across many domains and languages," says Kevin, "which is why I have worked on developing this new sentence segmentation engine."

TM-Town also released a test set of common edge cases that can be used to judge the accuracy of various segmentation tools. These so-called [Golden Rules](#) help to provide a standardized set of tests for various languages.

To learn more about TM-Town visit:
<https://www.tm-town.com>

About TM's for Translators, LLC:

TM's for Translators, LLC is a software development company dedicated to creating a better translation world through technology and specialization. With TM-Town, a patent-pending translation enablement platform, translators are empowered to reach relevant clients and jobs and manage and leverage their linguistic assets.

Contact

Kevin Dias
TM's for Translators, LLC
1-302-261-3852

