HOLLIS — **Segetes**

<0 images>

Segetes: A Framework for Creating and Curating Networked Texts

Hollis, L.

Reading on a screen should be as simple and beautiful of an experience as reading a book. Similarly, interacting with metadata should be as intuitive as interacting with the text itself, no matter what device a user accesses it with. Segetes is a simple, developer-friendly framework for creating and curating networked texts in web applications that can share seamlessly between each other and have the potential to create decentralized networks. The first application built with the Segetes framework includes the poetry of Vergil and is available at [segetes.io](http://segetes.io/).

Segetes’ project is twofold: first, to discover and extract metadata for source texts with its open-source modules—including grammatical structures, scansion, textual allusion, entity abstraction, related media, and a concordance of citations for line-specific related criticism—and second, to provide easy access to this data via a web interface, RESTful web service, or document exports as JSON-LD or XML. The Segetes framework is built to be as useful and streamlined as possible for developers at all skill levels and allows researchers to customize the data model and define schemas as necessary.

Segetes seeks to simplify user interactions and strives for clarity within intricacy in complex metadata. It accomplishes this by allowing users to toggle quickly between the most minimal view possible—the text only, without distractions—and the view for complex metadata related to each paragraph of prose or line of poetry. The primary endpoint of the API is the paragraph or line—and most facets of metadata reference this unit as a whole; thus, the data model provides ease of search and ordering by this base unit to make developing new views, applications, and components as easy as possible.

As networked texts continue to change the way that users read and study, Segetes seeks to make sharing data an integral and creative force that drives the formation of archives. Automated import and export modules remove obstacles between developers and their data and enable researchers to share it easily. These modules attempt to learn and predict features (such as section breaks and line numbers) about imported texts, and other modules for entity recognition, and calculating text reuse may be used to extract and aggregate other relevant metadata. The export data models rely on and extend current Schema.org, TEI, and Dublin Core standards for annotating metadata and utilize CITE architecture URNs where applicable to offer the most useful and sharable data model possible. Segetes recommends that applications built with its framework and modules contribute back to the community by publishing their datasets, whether as JSON-LD, XML, or another data format, on a larger collaborative data-sharing platform such as DataHub.

Similarly, each Segetes application offers robust webhooks for networking resources and updates between other Segetes applications that it is linked with. Text search and reuse, disambiguating shared entities, publishing updates, and sharing media are all made easy with in-network texts. In this way, applications developed and curated by individual researchers or teams of researchers and developers may network with each other and create decentralized archives of information, benefiting from sharing resources such as named entities and media, while distributing the burden of editing, data cleaning, hosting, and application development.



