

***Tycho Brahe Framework* – construa e administre seu próprio corpus**

(180 min)

Objetivos: Apresentar a plataforma de construção, administração e pesquisa de corpora linguísticos *Tycho Brahe Framework* e interagir com o público permitindo que eles mesmos possam, durante o workshop, criar um corpus com seus documentos e objetivos e possam utilizar todas as ferramentas disponibilizadas para catalogação, edição, recursos multimídias, etiquetamento, anotação sintática (parsing) e buscas.

Descrição: O *Tycho Brahe Framework* consiste em uma plataforma computacional web para administração de corpora linguísticos de diferentes naturezas incluindo línguas pouco estudadas como as polissintéticas, com recursos multimídias, tradução sentença a sentença, integração com etiquetadores morfológicos e parsers sintáticos, ferramentas gráficas para buscas morfossintáticas, resultados estatísticos, armazenamento de histórico de buscas, metadados, edição de árvores sintáticas dentro de um fluxo de trabalho, além de recursos para importação e exportação de dados nos formatos mais utilizados pelos pesquisadores. A plataforma se baseia na integração de microestruturas sintáticas armazenadas em uma base de dados exclusiva para informações sintáticas e em parâmetros computacionais que proporcionam uma grande flexibilidade na experiência do usuário.

O workshop será conduzido em duas partes. A primeira consiste na apresentação da plataforma *Tycho Brahe Framework* para os participantes mostrando como são utilizadas as ferramentas da plataforma e como são realizadas as configurações para diversos tipos de línguas (por exemplo, uma língua flexional como o português vs. uma língua polissintética (como o Kadiwéu), documentos (diálogos, cartas, textos de teatro, etc.) e recursos multimídia (tradução, som, imagem, localização, vídeo). Na segunda parte, os pesquisadores aprenderão a explorar a plataforma iniciando a construção e exploração de seus próprios corpora linguísticos (um por participante) através da supervisão dos instrutores.

Público-alvo: Pesquisadores, professores, estudantes de graduação e pós-graduação

É recomendável ao participante trazer seu computador pessoal e algum documento previamente selecionado para criação de corpus linguístico (textos, imagens, sons, etc. todos

em formato digital).

Instrutor: Luiz Veronesi - Possui graduação em Bacharelado em Ciência da Computação pela Universidade Federal de São Carlos (2002), graduação em Bacharelado em Letras pela Universidade de São Paulo (2009) e mestrado em Linguística pela Universidade de São Paulo (2015). Tem experiência na área de Ciência da Computação, com ênfase em análise e desenvolvimento de sistemas web. Responsável pela elaboração e desenvolvimento da plataforma *Tycho Brahe Framework* no âmbito do projeto temático de pesquisa da Fapesp sob coordenação da professora Charlotte Galves, UNICAMP. Atualmente trabalha como arquiteto de soluções em consultoria na área de Tecnologia da Informação.

Email: luiz.tycho@texugo.com.br

Websites de referências:

Corpus Tycho Brahe - <http://www.tycho.iel.unicamp.br>

Tycho Brahe Framework - <http://www.tycho.iel.unicamp.br/tbf>

Referências:

GALVES, C.; SANDALO, F.; SENA, T. A.; VERONESI, L. H. L. Annotating Polysynthetic Language: From Portuguese to Kadiwéu. *Cadernos de Estudos Linguísticos* – (59.3), Campinas, pp. 629-645. Set/Dez. 2017.

PAIXÃO DE SOUZA, M. C. "O Corpus Tycho Brahe: contribuições para as humanidades digitais no Brasil". *Filol. Linguíst. Port.*, São Paulo, v. 16, n. spe, p. 53-93, dez. 2014.

PAIXÃO DE SOUSA, M. C.; KEPLER, F. N.; FARIA, P. P. F. E-Dictor: Novas perspectivas na codificação e edição de corpora de textos históricos. In: Tania Shepherd; Tony Berber Sardinha; Marcia Veirano Pinto. (Org.). *Caminhos da linguística de corpus*. Campinas: Mercado de Letras, 2010.