ByteSpan: Information-Driven Subword Tokenisation

View arXiv paper here.