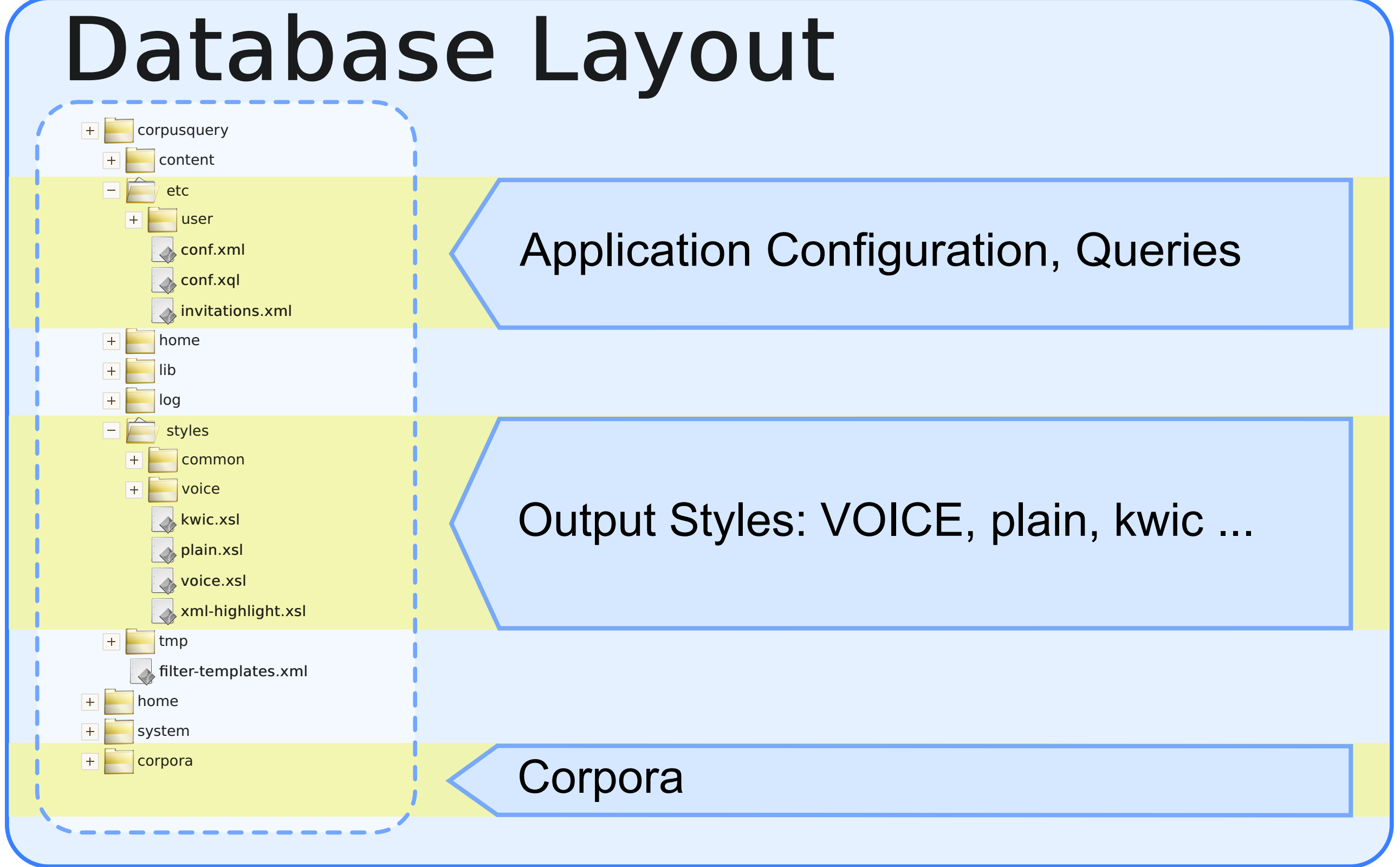
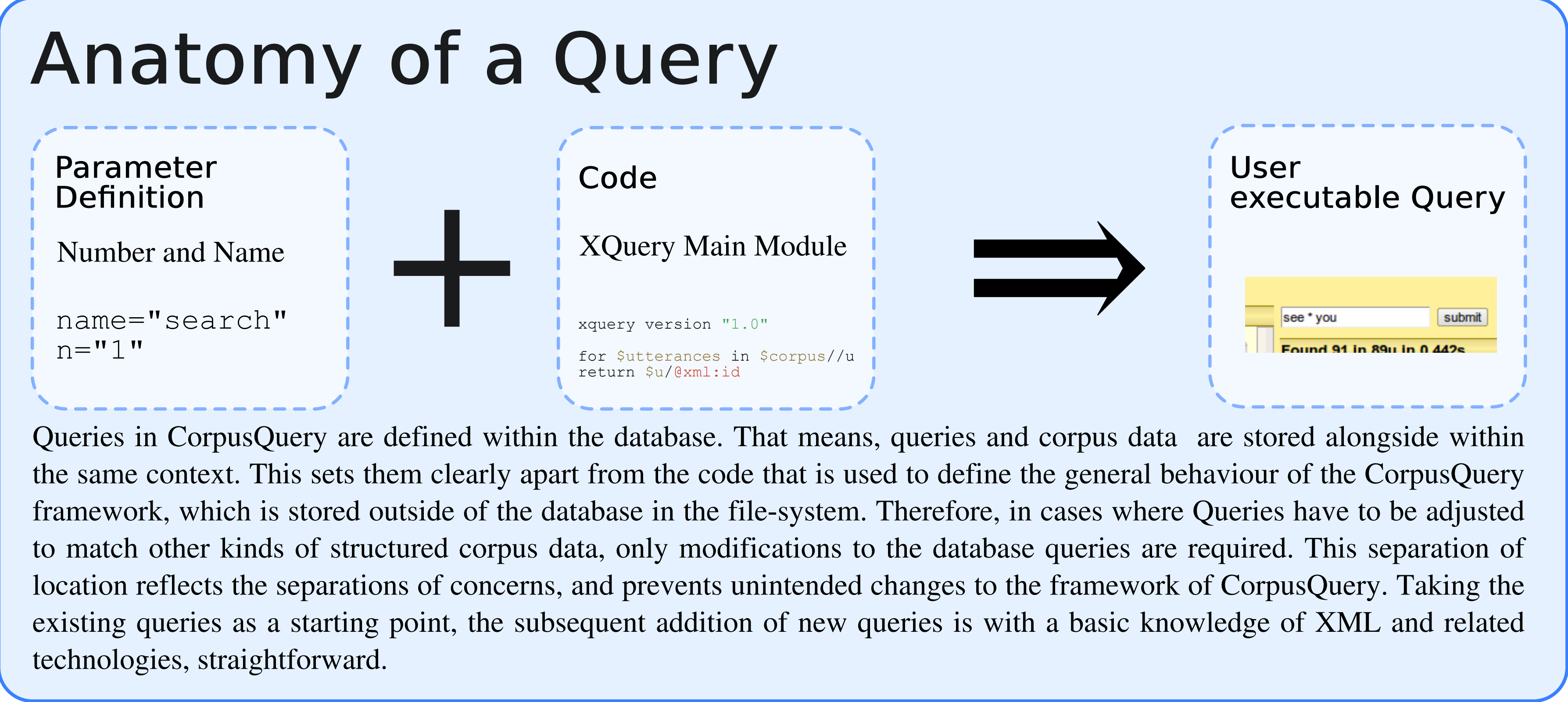
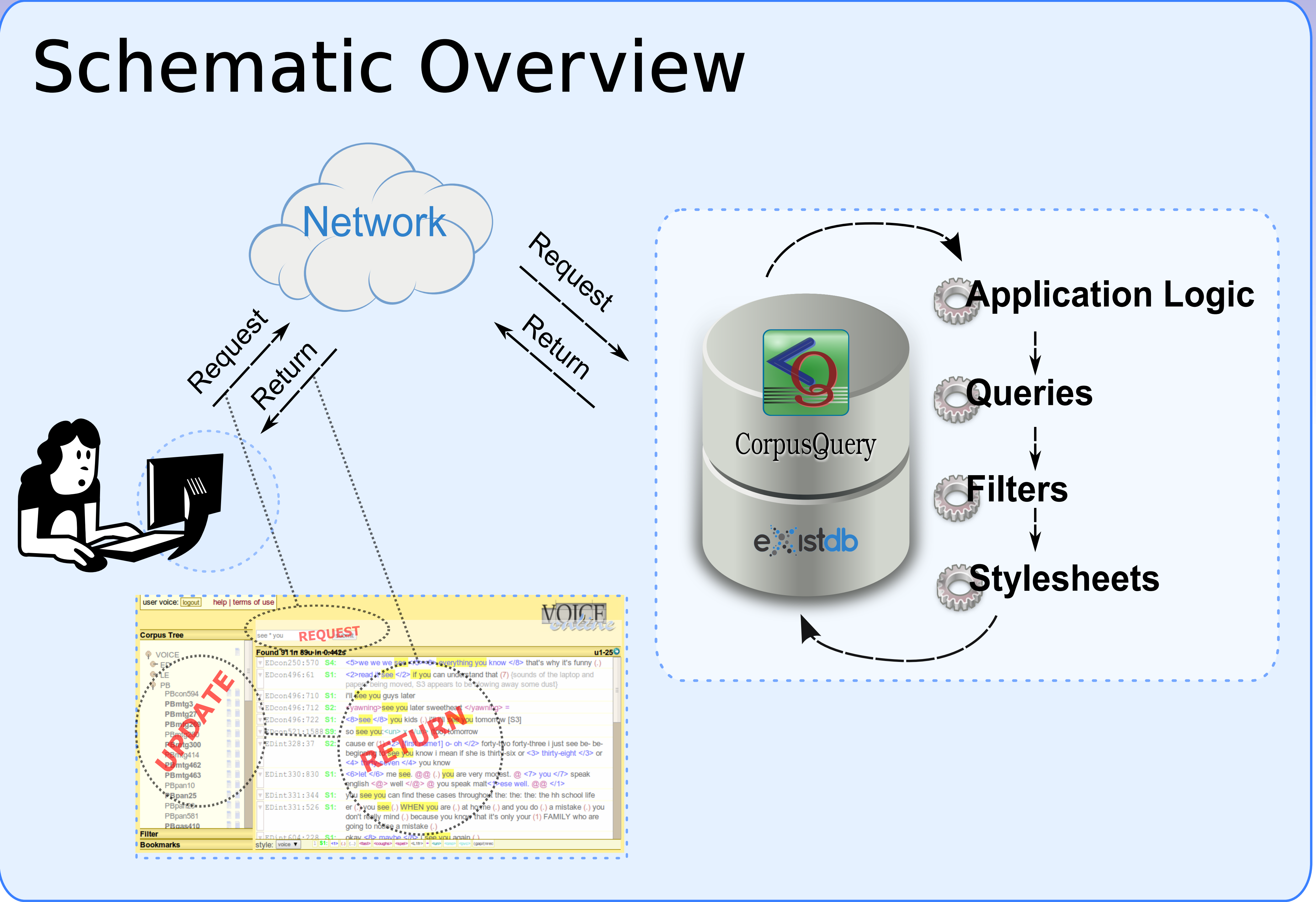
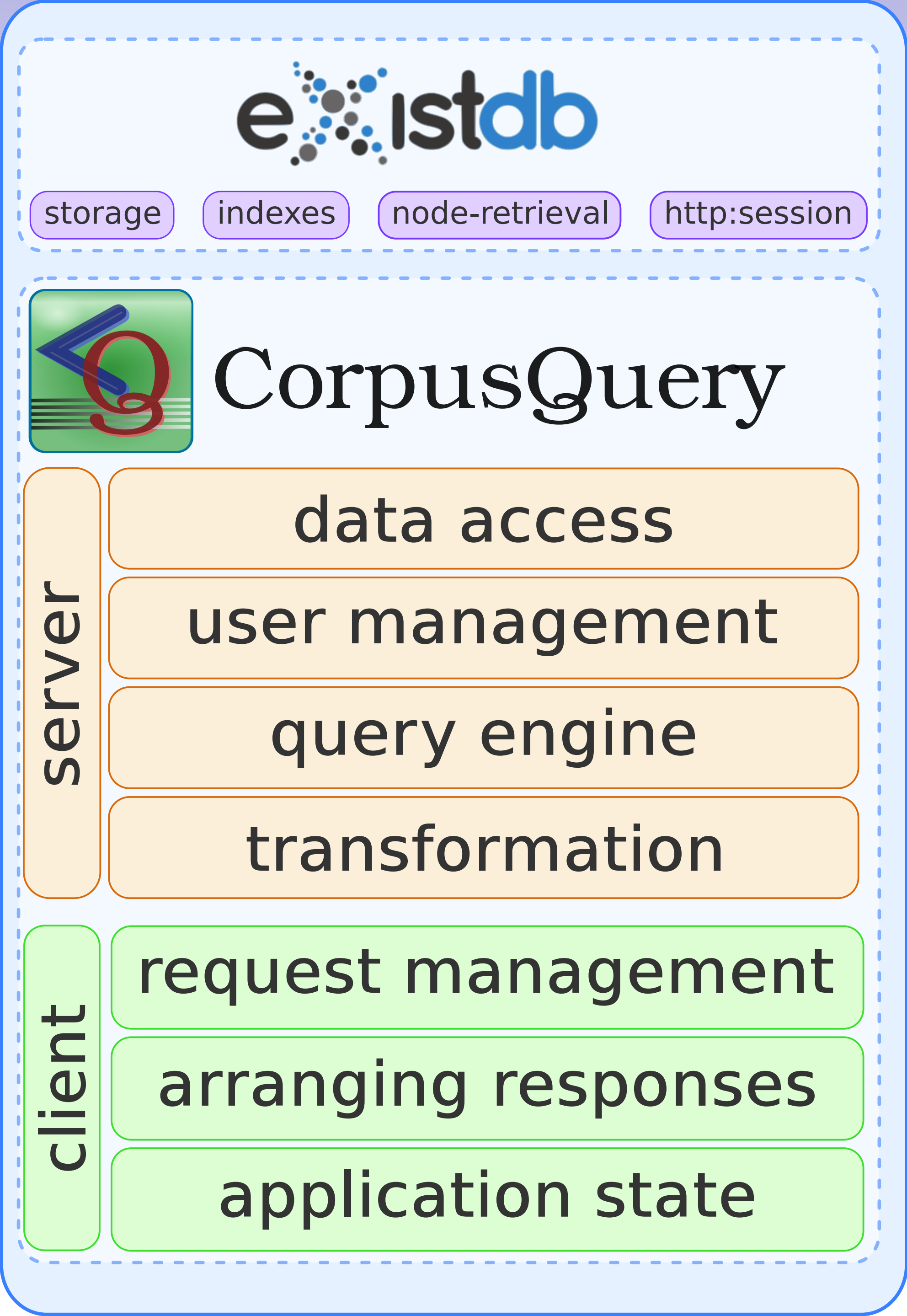
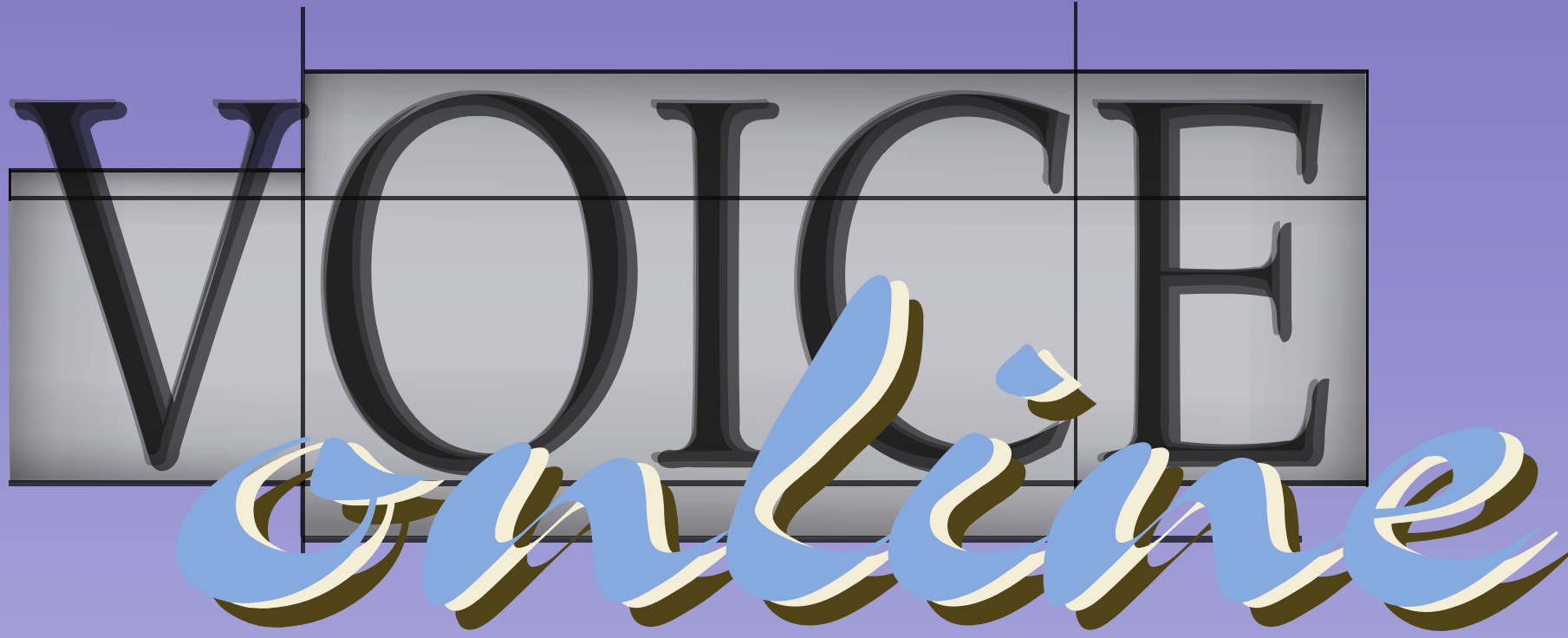




# CorpusQuery

Bringing VOICE to the web

Stefan Majewski, University of Vienna



**Abstract: VOICE Online, bringing VOICE to the web**

The Vienna-Oxford International Corpus of English (VOICE) contains transcriptions of interactive spoken communication featuring a rich set of meta-data. VOICE uses the Package CorpusQuery, which has been developed in the VOICE Project to make VOICE publicly accessible. CorpusQuery is an extensible search infrastructure for TEI-XML based corpora (cf. TEI Consortium 2007). Its current main purpose is the fast retrieval and presentation of pattern-based text searches within VOICE.

VOICE Online (cf. VOICE Project 2009) is a web-application built on the foundation of open XML standards using the native XML database eXist (cf. eXist-Project 2009). The general architecture is broken up into a server component, providing the search interface and a web-client for interactive search and retrieval of queries. The general design is abstract and modular. Therefore, a maintainer of a CorpusQuery installation, like VOICE Online, can customise the installation with only moderate effort and training. Hence, a maintainer can decide on the options for output styles and configure and define more query types that are publically available to the users.

The proposed visual presentation presents the rationale and architecture behind the design of CorpusQuery. It introduces users to the general principles and their application to the context of VOICE Online.

## Options for Extension

Following parts of CorpusQuery can be adapted to the specific needs of a particular CorpusQuery installation. For each modification, different standards have to be mastered. The fundamental requirement is a basic knowledge of XML and the ability to edit XML. The table below shows the basic options ordered by increasing difficulty.

Extension	Location	Requirement
Add Corpus texts	Database	copy /paste to network drive
Add Corpora	Database	above + basic XML editing
Add/Modify Queries	Database	above + basic XQuery
Add/Modify Styles	Database	above + basic XSLT
Change Application Framework	Filesystem	above + Javascript

## Select References

eXist-Project. 2009. eXist-db Open Source Native XML Database. <http://www.exist-db.org> (2009-11-26 14:44:59).

Seidlhofer, Barbara. 2005. "English as Lingua Franca". *ELT Journal* 59, 339-341

TEI Consortium. 2007. *Guidelines for Electronic Text Encoding and Interchange*. <http://www.tei-c.org/Guidelines/P5> (2009-11-19 23:00:00).

VOICE Project. 2009. *VOICE - Availability*. [http://www.univie.ac.at/voice/page/corpus\\_availability](http://www.univie.ac.at/voice/page/corpus_availability) (2009-11-26 14:27:03).

W3C. 2007. *XQuery 1.0: An XML Query Language*. <http://www.w3.org/TR/xquery> (2009-11-13 19:44:10).