



MPIWG

# Arboreal MWN

*Julia Damerow, Max-Planck-Institut für Wissenschaftsgeschichte*

## Inhaltsverzeichnis

Aufgabenbeschreibung .....	2
Funktionsanalyse.....	2
Quellcode Analyse.....	5
Workplan .....	6
Implementationsstatus .....	6

## Aufgabenbeschreibung

Allgemein ist das Ziel, die wesentlichen Funktionen des Programmes Arboreal, das ursprünglich von Malcolm Hyman entwickelt wurde und im Institut für die Erstellung von Transkriptionen und Übersetzungen altsprachlicher Texte benutzt wird, weiter zu entwickeln bzw. in andere Tools zu integrieren. Die letzte zur Verfügung stehende Version von Arboreal ist mit den aktuellen Versionen von Mac OSX nicht mehr kompatibel, so dass eine Weiterentwicklung dringend notwendig ist, außerdem ist nicht sichergestellt, dass der für die bestehenden Version notwendige externe Server für die Sprachtechnologien in Harvard mittelfristig weiter gepflegt wird, so dass ein Anbindung an den vom MPIWG betriebenen Sprachservices notwendig ist.

Zunächst wird aus der Ursprungsversion von Arboreal, die als ausführbare jar-Datei vorliegt, eine app-Datei erstellt werden, die auch auf neueren OSX Versionen (auf denen Rosetta nicht installiert ist) ausführbar ist. Im zweiten Schritt, wird der Quellcode überarbeitet und in das Framework Eclipse überführt werden. Die Reimplementierung von Arboreal im Eclipse Framework wird „Arboreal MWN“ genannt. Der nächste Abschnitt enthält die notwendige Funktionsanalyse.

## Funktionsanalyse

Im Folgenden werden die Funktionen der Ursprungsversion von Arboreal beschrieben und vermerkt ob und in welcher Art sie in Arboreal MWN übernommen werden sollen.

	Funktion		Beschreibung Unterfunktion	Aboreal	Arboreal MWN
1.	<i>Anzeigen von XML Dokumenten</i>	1.1	XML Dokumente werden in Baumansicht im linken Teil des Editorfensters dargestellt. Klicken auf einen bestimmten Knoten zeigt den diesen Knoten im rechten Teil des Editors an. Sollen mehreren Knoten angezeigt werden, müssen entsprechend mehrere Knoten selektiert werden.	✓	✓✓ Eventuell sollte es möglich sein, allen Inhalt eines Knotens anzuzeigen, ohne notwendigerweise alle Unterknoten selektieren zu müssen.
		1.2	XML kann auf Wunsch auch gerendert angezeigt werden.	✓	✓✓ Hierbei ist die Frage noch offen, ob Malcoms Code übernommen werden soll oder neuimplementiert.

<sup>1</sup> <https://it-dev.mpiwg-berlin.mpg.de/tracs/Arboreal>

<sup>2</sup> <http://www.eclipse.org/>

	Funktion		Beschreibung Unterfunktion	Aboreal	Arboreal MWN
		1.3	Baumansicht kann ohne XML angezeigt werden.	✓	✓
2.	<i>Bearbeiten von XML Dokumenten</i>	2.1	Der textuelle Inhalt von XML Dokumenten kann geändert werden.	✓	✓✓
		2.2	Der strukturelle Inhalt von XML Dokumenten kann geändert werden		✓ (Mögliche Erweiterung)
3.	<i>Cloning</i>	3.1	Arboreal ermöglicht es, die Struktur von Dokumenten zu klonen. Dabei wird die XML Struktur des Ausgangstextes übernommen, jedoch wird der Inhalt und alle IDs gelöscht.	✓	✓✓  Dies könnte erweitert werden, indem dem Nutzer die Möglichkeit gegeben wird, die IDs des Ausgangstextes in den Klon zu übernehmen oder neue IDs zu erstellen und auf Wunsch ein Matching-file automatisch von Ausgangstext und Klon zu erstellen.
4.	<i>Dokumenten Matching</i>	4.1	Gibt es zu mehreren XML Dokumenten Matching files (Dateien, die den Elementen IDs der einen XML-Datei, IDs der anderen XML-Datei zuordnen), können diese Dokumente parallel dargestellt werden (m-zu-n Kollationen). Dabei können die "gematchten" Dokumente parallel mitbewegt werden, beim Durchgehen einer der Texte.	✓	✓✓

	Funktion		Beschreibung Unterfunktion	Aboreal	Arboreal MWN
		4.2	Erstellen von Matching files		✓ Bisher bietet Arboreal keine Möglichkeit Matching files einfach zu erstellen. Dies wäre eine mögliche Erweiterung (z.B. graphisch)
5.	<i>Suche</i>	5.1	Es kann mit RegEx und (wahlweise) orthographischer und morphologischer Normalisierung gesucht werden.	✓	✓✓
		5.2	Es gibt zwei mögliche Suchergebnisanzeigen: - als (parallele) Liste mit Container, umgebendem Text, IDs, etc. - im Text	✓	✓
		5.3	Bei paralleler Darstellung von mehreren Files (mit Matching files) bewegen sich Paralleltexte mit wenn durch man sich durch die Suchergebnisse im Haupttext bewegt.	✓	✓
6.	<i>Morphologische Analyse</i>	6.1	Analyse kann heruntergeladen werden, so dass danach offline gearbeitet werden kann.	✓	✓✓
		6.2	Ergebnisse der morphologischen Analyse können für Wörter angezeigt werden.	✓ Momentan im Contextmenu	✓✓ Ergebnisse sollten in Properties-View oder ähnliches angezeigt werden.
		6.3	Morphologische Analyse nicht mehr über Donatus, sonder by Default über Josephs Tool. Sollte konfigurierbar sein.		✓✓
7.	<i>Termlisten</i>	7.1	Termlisten können angelegt werden (mit Erläuterungen, Normierungen und Anbindung an Text)	✓	✓

	Funktion		Beschreibung Unterfunktion	Aboreal	Arboreal MWN
		7.2	Liste aller Instanzen der Terme	✓	✓
		7.3	Angabe Anzahl der Instanzen im Text zum Vergleich im ganzen Korpus.	✓	✓
		7.4	Klick auf Term in Termliste zeigt liste aller Instanzen und/oder Stelle im Text.	✓	✓
		7.5	Bei Texten mit mehreren Termlisten werden alle Terme einer Liste in eigener Farbe angezeigt.		✓
		7.6	Verlinken nach ECHO	✓	(✓) Bei Anbindung an Repository, sollte man zu diesem verlinken können.
		7.7	Anbindung an Text: bisher müssen Anzahl der Termteile übereinstimmen für Anbindung, sollte geändert werden so dass "Schwerkraft" auch an "Kraft der Schwere" angebunden werden kann.		✓
8.	<i>Repository</i>	8.1	Texte können direkt aus Repository geholt werden.		✓
9.	<i>Projekte</i>	9.1	Texte werden über Projekte verwaltet. Ein Projekt enthält alle benötigten Dateien.		✓✓
10.	<i>Rendering</i>	10.1	DocSpecs ersetzen. Andere Lösung für DocSpecs Mechanismus finden.		(✓)
11.	<i>Sprachen</i>	11.1	Chinesisch/Japanisch/... Morphologische Analyse, wo es keine Blanks als Wortgrenzen gibt.		(✓)

## Quellcode Analyse

Das Framework Eclipse nutzt eine andere Java-Bibliothek zur Erstellung graphischer Benutzeroberflächen (SWT/JFace) als die Ursprungsimplementation Arboreals (AWT/Swing). Die Integration der AWT/Swing-Oberfläche würde ähnlich viel Zeit in Anspruch nehmen wie die Neuimplementierung in SWT/JFace, daher wird die Benutzeroberfläche von Arboreal

MWN in SWT und JFace erneut implementiert werden. Dies erlaubt überdies eine einfache Integration von Arboreal in TextGrid.

Das Backend von Arboreal, beispielsweise die morphologische Analyse oder der Suche, sollte zumindest teilweise übernommen und angepasst werden and Arboreal MWN. Allerdings wird die Anbindung und Konfigurierbarkeit an andere Sprachservices ebenfalls eine Neuimplementation nötig machen.

## Workplan

Zunächst werden die Kernfunktionen von Arboreal implementiert werden. Kernfunktionen sind in der Spalte (Arboreal MWN) mit zwei Haken gekennzeichnet (✓✓). Danach werden alle Funktionen, die mit einem Haken gekennzeichnet sind, umgesetzt (✓). Bei genügend Zeit werden dann die Funktionen mit geringster Priorität, gekennzeichnet durch einen eingeklammerten Haken, implementiert.

## Implementationsstatus

	Funktion		Unterfunktion	Status
	<i>Erstellung einer auf OSX Lion lauffähigen Version von Arboreal</i>			✓
1.	<i>Anzeigen von XML Dokumenten</i>	1.1	Baumansicht von XML-Dokumenten	✓
2	<i>Bearbeiten von XML Dokumenten</i>	2.1, 2.2	Inhalt und Struktur kann über Eclipse XML Editor geändert werden.	✓
6.	<i>Morphologische Analyse</i>	6.1 - 6.3	Anschluss an Josephs Tool; Analyse wird heruntergeladen und kann für analysierte Wörter angezeigt werden.	✓
7	<i>Termlisten</i>	7.1,, (7.2)	Termlisten können erstellt werden. Instanzen werden im XML festgehalten.	✓
9.	<i>Projekt</i>	9.1	Die Projektverwaltung von Eclipse wurde in Arboreal integriert. Texte werden über Projekte verwaltet.	✓