

Konzept für das Redaktionssystem des DAG

15. Dezember 2014

Zusammenfassung

Dieses Dokument beschreibt ein Konzept für die Entwicklung eines elektronisch gestützten Redaktionssystems für das Wörterbuch *Dictionnaire d'ancien gascon* – DAG, erarbeitet bei der Heidelberger Akademie der Wissenschaften. Dieses Redaktionssystem wird die Redaktion von sogenannten Kurzartikel des DAG, DAGpré-Artikel, ermöglichen. Die Grundlage der Entwicklung des DAG-Redaktionssystems wird das Redaktionssystem des *Dictionnaire étymologique de l'ancien français* – DEAF sein. Dieses Dokument beschreibt, was das zu entwickelnde System leisten soll und wie es eingesetzt werden wird.

Inhaltsverzeichnis

1. Zielsetzung	6
2. Vorarbeiten	8
2.1. Benutzerverwaltung	8
2.2. Zettleingabe	8
2.3. Bibliographie	9
3. Funktionale Anforderungen	11
3.1. Hauptmenü	11
3.1.1. Verwaltung der XML-Elemente	12
3.1.2. Zettel-Lemmata-Abgleich	12
3.1.3. Liste POS	12
3.1.4. Sigel: Ansicht und Bearbeitung der Bibliographie	12
3.1.5. HW-Kategorien	12
3.1.6. Zettelkästen	12
3.1.7. XML-Export	13
3.2. Lemmatisierung	13
3.2.1. Zettel-Lemmata-Abgleich	13
3.3. Erstellung von Kurzartikeln	13
3.3.1. Der „Titelblock“	14
3.3.2. Der „Wörterbuchblock“	14
3.3.2.1. Anzeige des Wörterbuchblocks	14
3.3.2.2. Etymologie	15
3.3.3. Der „Graphienblock“	15
3.3.4. Der „Bedeutungsblock“	16
3.3.4.1. Festlegung der Bedeutungsstruktur	16
3.3.4.2. Zuordnen der Zettel zu Bedeutungen	17
3.3.5. Der „Kommentarblock“	18
3.3.6. Semantisches Markup in Freitextfeldern	18
3.4. Datenkodierung	19
3.5. Verknüpfung mit externen Datenbeständen	19
3.6. Elektronische Publikation	19
3.6.1. Suchfunktionen	20
A. Überblick Systemarchitektur	21
B. Datumsangaben	22
C. Part-of-Speech Tags	23
D. XML-Elementeliste	24

Inhaltsverzeichnis

E. Hauptmenü	25
F. Zetteleingabemaske	26
G. Eingabemaske der Bibliographie	27

1

Zielsetzung

Die Entwicklung des DAG-Redaktionssystems ist Teil des Projekts „Digitalisierung des DAG“ – DigiDAG. Weitere Module des Projektes sind die Digitalisierung des Zettelkastens des DAG (mit ca. 200.000 Zetteln), die Digitalisierung der Bibliographie des DAG und schließlich die Verknüpfung der Datenbank des DAG mit der anderer Projekte zum (Alt-)Gaskognischen und (Alt-)Okzitanischen. Einen Überblick über die Systemarchitektur gibt die Abbildung in Anhang A.

Für die Konzipierung des Redaktionssystems wird hier die Digitalisierung des Zettelkastens sowie die der Bibliographie als gegeben angenommen, obwohl diese Module zum Zeitpunkt der Konzepterstellung noch nicht abgeschlossen sind. Die Verknüpfung der Datenbank des DAG mit der anderer Projekte ist ein zukünftiges Modul und nicht Inhalt dieses Konzepts.

Nach Abschluss aller Module kann die Redaktion der DAGpré-Artikel erfolgen. Das Resultat soll in einer elektronischen Publikation mit *open access* zur Verfügung gestellt werden: „DAG électronique“ – DAGél. DAGél soll aus der Bibliographie (DAGBiblél), DAGpré und evtl. aus einer retrodigitalisierten Version der bereits als Buch publizierten Faszikel des DAG bestehen.

Das Redaktionssystem soll im einzelnen Aspekte der

- Datenverwaltung (Zettel-, Bibliographie-, Artikeldaten)
- Benutzerverwaltung (An-/Abmeldung, Rechtezuordnung, Datenzuordnung, etc.)
- Prozessverwaltung (definierter Workflow der Artikelerstellung)
- Kontextsensitive, semantische Unterstützung (Anzeige relevanter Informationen bei der Artikelerstellung)
- XML-Export der Daten zur Archivierung
- XHTML-Export der Daten zur elektronischen Publikation

innerhalb eines integrierten Informationssystems implementieren.

Soweit wie möglich werden bei der Entwicklung das Redaktionssystem die Strukturen des Redaktionssystems des DEAF eingesetzt und diese werden, wo nötig, an die Bedürfnisse der Redaktion von DAGpré-Artikeln angepasst. Unterschiede zwischen den beiden Wörterbüchern DAG und DEAF machen eine Adaption des DEAF-Redaktionssystems unabdingbar.

Das Ziel bei der Entwicklung ist, die Redaktion von technischen Details der Informationsverwaltung und -verarbeitung zu entlasten, damit diese sich auf ihre Kernaufgabe, die Forschung und Entwicklung am Wörterbuch, widmen kann. Ein besonderer Aspekt ist die Tatsache, dass das DAG-Projekt über keine eigene EDV-Kraft verfügt. Dies muss bei der Erarbeitung der Installations-, Betriebs-, und Wartungskonzepte berücksichtigt werden. Wir weisen hier darauf hin, dass es nicht möglich sein wird, ein solches elektronisch gestütztes Redaktionssystem über Jahre hinweg zu benutzen, ohne Wartungsarbeiten auszuführen.

2

Vorarbeiten

Die folgenden Vorarbeiten sind im Rahmen des Projekts „DigiDag“ bereits geleistet worden, bzw. befinden sich in Arbeit.

2.1. Benutzerverwaltung

Das System arbeitet sitzungsorientiert, jedem Benutzer wird dazu eine Benutzerkennung und ein Kennwort zugeordnet, die zur Identifikation gegenüber dem System dienen. Die folgenden Funktionen werden benötigt:

Anmeldung: vor dem Zugriff auf die Daten des Systems ist eine Anmeldung mit Benutzernamen und Kennwort erforderlich.

Benutzerverwaltung: ein Benutzer mit Administratorrechten hat die Möglichkeit zum Anlegen neuer Benutzer oder zum Deaktivieren existierender Benutzer. Deaktivierte Benutzer können sich nicht mehr anmelden.

2.2. Zetteleingabe

Der Zettelkasten des DAG enthält ca. 200.000 Zettel, die digitalisiert werden müssen. Zu diesem Zweck ist eine Zetteleingabemaske entwickelt worden. Sie ist in die bereits angelegte Benutzeroberfläche integriert.

Die folgende Tabelle beschreibt die Eingabefelder der Zetteleingabemaske. Die eingegebenen Informationen werden in Datenbankfeldern gespeichert. Die XML-Elemente sind bereits im Hinblick auf den zukünftigen XML-, bzw. XHTML-Export für die elektronische Publikation festgelegt.

	XML-Element	Bezeichnung in Maske	Kommentar	Bsp.
1	entry	Wort	Form, so wie auf Zettel	<i>retorns</i>
2	lemma	Lemma	unflektierte Form [so oft nicht auf Zettel]	<i>retorn</i>
3	pos	Wortart	oft nicht auf Zettel	<i>s.m.</i>
4	siglum	Sigel	automatische Eingabekorr. mit DAGBibl	ACoutBay
5	date	Textdatierung	aus DAGBibl	1402 (n.st)
6	singular-date	Kopiedatum	Einzel-, Kopiedatum innerhalb einer Textausgabe	ca. 1400
7	loc	Lokalisierung	Lokalisierung des Quelle	Bayonne
8	text-reference	Textstelle	Stellenverweis	88 - VI 199
9	language	Sprache	Sprache der Quelle	agasc. [default]
10	ref-onomas	Kategorie	Hallig-Wartburg-Kategorie	B III b 6
11	onomasiology	Begriff	Konzept zur HW-Kategorie	<i>propriété</i>
12	definition	Definition	Bedeutung des Wortes	“navire”
13	enc-def	Enzykl. Definition	längere Definition, enthält Enzyklopädisches, etc. (oft nicht auf Zettel)	“petit bateau terminé en pointe par les deux bouts”
14	context	Kontext	Kontext des Wortes	<i>Cum filha deu aver retorn en so que lo payre bent</i>

S. den Screenshot im Anhang **F**.

2.3. Bibliographie

Zum Zeitpunkt der Konzepterstellung sind die bibliographischen Einträge zu den Sigeln des DAG in einer odt-Datei festgehalten.

Im Rahmen von DigiDAG wurde bereits eine Digitalisierung der Bibliographie konzipiert und die benötigte Eingabemaske wurde entwickelt. Sie dient dazu, die Informationen der bibliographischen Einträge mit Markup auszuzeichnen und die Datenbankrelevanten Informationen zu markieren (s. Anhang **G**).

Die Digitalisierung der Bibliographie ist die Grundlage für die folgenden drei Funktionen:

1. Die in Datenbankfeldern abgelegten Informationen dienen zum automatischen Erkennen und gegebenenfalls Korrigieren der eingegebenen Sigel während der Zetteleingabe. Anhand des Abgleichs der eingegebenen Zeichenketten mit den

2. Vorarbeiten

Informationen aus der Datenbank wird verhindert, dass falsche Sigel abgespeichert werden. Dadurch soll ein nachträglicher Korrekturvorgang der Sigel unnötig werden.

2. Die Informationen Sigel + Sigeldatum dienen während der Redaktion der DAG-Kurzartikel zur automatischen chronologischen Sortierung der Sigel, die in den Tabellen von „Wörterbuchblock“ (s. Abschnitt 3.3.2), „Graphienblock“ (s. Abschnitt 3.3.3) und „Bedeutungsblock“ (s. Abschnitt 3.3.4) angezeigt werden.
3. Die digitale Bibliographie soll elektronisch publiziert werden. Anhand einer Verknüpfung der in den DAG-Kurzartikeln zitierten Sigel mit den entsprechenden Sigelinträgen der Bibliographie soll dem Nutzer von DAGél die unmittelbare Auflösung der Sigel ermöglicht werden.

Es ist geplant, Mitte Dezember 2014 mit der Digitalisierung der Bibliographie zu beginnen.

Ob die bibliographischen Einträge eine Zuordnung zu Primär-, Sekundär- und Tertiärliteratur erhalten, ist zum Zeitpunkt der Konzepterstellung noch ungeklärt. S. dazu im Abschnitt 3.3.2 zu den 'Unterschieden zu den Funktionen des DEAF-Redaktionssystems'.

3

Funktionale Anforderungen

Es werden die funktionalen Anforderungen an alle Aspekte der elektronisch gestützten Redaktion der Artikel (Hauptmenüführung und Kurzaufartikelredaktion) beschrieben.

Die Verknüpfung mit externen Datenbeständen sowie die elektronische Publikation von „DAG électronique“ werden angesprochen, sind jedoch im Detail nicht Teil des vorliegenden Konzepts. S. dazu die Abschnitte 3.5 und 3.6.

3.1. Hauptmenü

Nach der Anmeldung kann aus den folgenden Menüpunkten ausgewählt werden:

1. Artikelauswahl
Bearbeiten von Kurzaufartikeln aus einer gegebenen Liste von Artikeln (s. Abschnitt 3.3)
2. Datenpflege
 - a) Verwaltung der XML-Elemente (s. Abschnitt 3.1.1)
 - b) Funktion „Zettel-Lemmata-Ableich“ (s. Abschnitt 3.2 und 3.2.1)
 - c) Ansicht einer gegebenen Liste von POS (*part of speech* / Wortarten, s. Abschnitt 3.1.3)
 - d) Zettelkästen: Eingabe und Bearbeitung der Zettel (s. Abschnitt 3.1.6)
 - e) Sigel: Ansicht und Bearbeitung der Bibliographie (s. Abschnitt 3.1.4)
 - f) HW-Kategorien: Ansicht und Bearbeitung der Hallig-Wartburg-Kategorien
3. Verwaltung
 - a) Benutzerverwaltung (s.o. Abschnitt 2.1)

3. Funktionale Anforderungen

Im Anhang (E) befindet sich ein Screenshot des Hauptmenüs (zum Zeitpunkt der Projekterstellung) mit den bereits implementierten Funktionen, die für die Zetteleingabe und für die Digitalisierung der Bibliographie verwendet werden.¹

3.1.1. Verwaltung der XML-Elemente

Über das Hauptmenü hat der Benutzer Zugang zur Verwaltung der XML-Elemente, die in der Liste des semantischen Markups erscheinen (s. 3.3.6). Der Benutzer kann XML-Elemente hinzufügen, Attribute zu einem bestehenden XML-Element hinzufügen und die möglichen Inhalte der Attribute bearbeiten (hinzufügen, löschen).²

3.1.2. Zettel-Lemmata-Abgleich

Zur Funktion „Zettel-Lemmata-Abgleich“, s. Abschnitt 3.2 und 3.2.1.

3.1.3. Liste POS

Die Ansicht einer gegebenen Liste von POS-Angaben (*part of speech* / Wortarten) kann über das Hauptmenü geöffnet werden. In ihr ist es möglich, POS-Angaben zu korrigieren, zu löschen und neue hinzuzufügen. S. Anhang 3.1.3.

3.1.4. Sigel: Ansicht und Bearbeitung der Bibliographie

Über das Hauptmenü können die Eingabemaske der Bibliographie aufgerufen werden, neue Sigel eingegeben und Sigel korrigiert werden. S. dazu den Screenshot im Anhang G.

3.1.5. HW-Kategorien

Im Rahmen von DigiDAG wird das Onomasiologische Begriffssystem von R. Hallig und W. v. Wartburg³ digitalisiert werden müssen. Nach den von Hallig und Wartburg festgelegten Kategorien sind der Zettelkasten und auch die bereits publizierten Artikel des DAG geordnet. Die Digitalisierung der HW-Kategorien dient zwei Zwecken: a) zum Verhindern von fehlerhafter Eingabe dieser Kategorien während der Zetteldigitalisierung und b) zum automatisierten Einfügen dieser Kategorien bei der Redaktion des „Bedeutungsblocks“ der DAGprä-Artikel, s. Abschnitt 3.3.4.

3.1.6. Zettelkästen

Über das Hauptmenü gelangt der Benutzer in die Bearbeitung der Zettel. Zum einen kann hier die Zetteleingabemaske geöffnet werden, zum anderen eine Tabellenansicht der Gesamtheit der digitalisierten Zettel. In dieser Tabelle ist das Editieren und Korrigieren sowie das Löschen von Zetteln aus dem System möglich. S. Anhang F.

¹Konzept Sabine Tittel / Marcus Husar; Entwicklung Marcus Husar.

²Zum gegenwärtigen Zeitpunkt ist es – im Redaktionssystem des DEAF – nicht möglich, die Scripte für den XML-, und den XHTML-Export (und den \LaTeX -Export) automatisch den Änderungen im System anzupassen. S. Abschnitt 3.6.

³R. Hallig – W. von Wartburg, Begriffssystem als Grundlage für die Lexikographie / *Système raisonné des concepts pour servir de base à la lexicographie*, 2^e éd., Berlin (Akademie-Verlag) 1963 (DEAF-Sigel: HW).

3.1.7. XML-Export

Unabhängig von der internen Datenstruktur des Systems müssen die DAGpré-Artikel in einem XML-Format exportiert werden. Die XML-Struktur der Daten muss in einer Schemadatei beschrieben werden. Dies ist nicht Teil des vorliegenden Konzepts.

Ein Rückimport (Übernahme extern geänderter XML-Dateien in das Redaktionssystem) ist nicht vorgesehen.

3.2. Lemmatisierung

Die digitalisierten Zettel enthalten als zentrale Information ein „Wort“ des Altgaskognischen. Die „Wörter“ müssen in einem ersten Arbeitsschritt lemmatisiert werden, da sie z.B. als flektierte Formen eines Verbs vorkommen können oder in unterschiedlichen graphischen Varianten belegt sein können. Durch das Lemmatisieren wird jedes „Wort“ mit den dazugehörigen Zettelinformationen eindeutig einem Lemma zugeordnet. Die Liste der Lemmata ergibt die Liste der zu redigierenden Kurzartikel.

z.B. *complir* und *cumplir*

Jedes Lemma ist mit den Zetteln seiner „Wörter“ verknüpft. Über diese Verknüpfung lassen sich sowohl ausgehend von einem Zettel das zugeordnete Lemma, sowie von einem Lemma alle Zettel erreichen.

Bsp. werden die Wörter (Zettel) *aparelhat* und *apareilhat* beide auf das Lemma *apareilad* abgebildet.

Der Schritt des Lemmatisierens könnte auf zwei Arten erfolgen:

1. Bei der Digitalisierung der Zettel wird jedes „Wort“ einem Lemma zugeordnet. Die Zetteleingabemaske enthält das entsprechende Eingabefeld.
2. In einem eigenen Arbeitsschritt werden die Zettel mithilfe des „Lemmings“ des DEAF lemmatisiert. Der „Lemming“ müsste dafür angepasst werden (z.B. benötigt das System des DAG keine Unterscheidung in Hauptlemma und Unterlemma; die halbautomatische altfranzösische Lemmatisierung mithilfe des Tobler-Lommatzsch könnte gelöscht werden). Er würde eine zeitsparende Möglichkeit bieten, viele Zettel auszuwählen und gleichzeitig einem Lemma zuzuordnen. Der Lemming ist ein separat implementiertes Tool. Er lädt Daten aus der Datenbank lokal auf den Arbeitsplatzrechner, bearbeitet diese Daten lokal und synchronisiert im Anschluss die Daten wieder mit der Datenbank.

Welche Art der Lemmatisierung angestrebt werden wird, war zum Zeitpunkt der Konzepterstellung noch nicht festgelegt.

3.2.1. Zettel-Lemmata-Abgleich

Im Falle des Verwendens des Lemmings müsste die Funktion des Zettel-Lemmata-Abgleichs vom DEAF-System übernommen werden. Sie dient dazu, die außerhalb der Datenbank lemmatisierten Daten mit denen der Datenbank zu synchronisieren.

3.3. Erstellung von Kurzartikeln

Aus einer gegebenen Liste der Kurzartikel, die sich aus dem Prozess des Lemmatisierens ergibt, wird ein Kurzartikel zur Bearbeitung ausgewählt und geöffnet. Die Redaktion des Kurzartikels ist in vier Blöcke aufgeteilt. Jeder der nachfolgenden Unterabschnitte repräsentiert den Bearbeitungsabschnitt eines Blockes und die daraus resultierenden Anforderungen.

3. Funktionale Anforderungen

3.3.1. Der „Titelblock“

In diesem Teil der Artikelerstellung wird der Artikelkopf bearbeitet. Dieser besteht aus dem Titelwort (= Lemma), der Angabe der Wortart (POS) und evtl. einer Sprach- bzw. Dialektangabe. Als Ergebnis dieses Schrittes entsteht der Titelblock des Kurzartikels. Soll ein Titelwort nachträglich zur Lemmatisierung der Zettel geändert werden, so kann dafür eine separate Oberfläche geöffnet werden. Diese erlaubt es, das Lemma zeitgleich für alle ihm anhängenden Zettel zu ändern.

Die Bearbeitung umfasst die folgenden Aspekte:

- (evtl.) eine Möglichkeit zu markieren, dass es sich um eine erschlossene Lemmaform handelt
- Auswahl der Wortarten aus einer Liste von Wortarten
- (evtl.) Auswahl einer Sprache/eines Dialektes aus einer Liste von Sprachen/Dialekten.

Beide Listen (die der Wortarten (POS-Tags) und die der Sprachen) können aus beliebig vielen Elementen bestehen, und sie müssen mindestens je ein Element enthalten. Bei der Auswahl der Wortarten bzw. der Sprachen/Dialekte können Elemente aus der Menge aller Wortarten (s. Anhang C) bzw. der Menge aller Sprachen/Dialekte ausgewählt werden; es können Elemente gelöscht, sowie Elemente hinzugefügt werden.

Bsp. *adj.* für Adjektiv;
béarn. für Bearnesisch

Unterschiede zu den Funktionen des DEAF-Redaktionssystems

Die Erarbeitung des Titelblocks kann möglicherweise erhalten werden. Die Liste der Wortarten muss evtl., die der Sprachen/Dialekte muss sicher dem DAG angepasst werden.

3.3.2. Der „Wörterbuchblock“

In diesem Schritt wird der Wörterbuchblock des Kurzartikels bearbeitet. Dieser besteht aus einer Liste von Wörterbuchverweisen und von Sekundärliteraturverweisen, die ein Ergebnis des Digitalisierens der Zettel ist.

3.3.2.1. Anzeige des Wörterbuchblocks

Im Wörterbuchblock stehen Verweise auf Sekundär- und Tertiärliteratur, in der das Artikelwort gefunden werden kann.

Die Festlegung des Wörterbuchblocks geschieht in mehreren Schritten:

1. Das System selektiert zunächst alle Zettel, deren Lemma mit dem Titelwort identisch ist. Zu jedem Zettel wird anhand der Bibliographiedatenbank nachgeschlagen, ob es sich um eine Sekundär- oder Tertiärquelle handelt.
2. Es werden nun zwei Listen angezeigt:
 - a) Liste der Tertiärliteratur, sortiert nach einer vorgegebenen Ordnung.
 - b) Liste mit Sekundärliteratur, sortiert nach einer vorgegebenen Ordnung oder alphabetisch nach dem Sigel.

Beide Listen können nun, unabhängig voneinander, manuell umsortiert werden, um die endgültige Reihenfolge für den aktuell bearbeiteten Artikel festzulegen. Diese Umsortierung wirkt sich nicht auf andere Artikel mit den gleichen Sigeln aus. Dabei können Zettel auch aus den beiden Listen entfernt werden, sprich aus dem System gelöscht werden.

3.3.2.2. Etymologie

Es muss noch spezifiziert werden, ob nach Vorbild des DEAF auch eine Art Kurz-etymologie auf Basis der Zettel des *Französischen Etymologischen Wörterbuchs* – FEW eingefügt werden soll. Die Etymologie kann als freier Text in einem Textfeld mit Markup-Unterstützung (wie in Abschnitt 3.3.6 beschrieben) eingegeben werden und wird so dem Kurzartikel zugeordnet.

Das System unterstützt diesen Schritt, indem es aus der Menge der Zettel, die dem Titelwort zugeordnet sind, alle Zettel mit dem Sigel „FEW“ auswählt. Von diesen werden das Sigel (Tag `siglum`), die Referenzangabe (Tag `reference`) und die Etymologie (Tag `cited-word`) zu Beginn in das Textfeld eingefügt und mit entsprechenden Markup umgeben. Dort kann, wenn nötig, die Information manuell nachbearbeitet oder gelöscht werden.

Unterschiede zu den Funktionen des DEAF-Redaktionssystems

Generell muss der folgende Unterschied berücksichtigt werden, der die drei Blöcke „Wörterbuchblock“, „Graphienblock“ (s. Abschnitt 3.3.3) und „Bedeutungsblock“ (s. Abschnitt 3.3.4) betrifft: Die Zettel des DAG sind in der Menge und Art ihrer Informationen und somit in ihrer digitalen Struktur nicht mit denen des DEAF deckungsgleich. D.h., dass überall dort, wo Zettel in einer Tabelle angezeigt werden, diese Tabellen den Bedürfnissen des DAG angepasst werden müssen.

Ein Aspekt, der für den Wörterbuchblock relevant ist, ist die Unterscheidung der Zettel in Primär-, Sekundär- und Tertiärzettel wie beim DEAF. Falls diese Unterscheidung nicht beibehalten werden kann (sie ist abhängig von der redaktionellen Nachbearbeitung der Bibliographie) muss der gesamte Wörterbuchblock überdacht werden.

Die Frage zum Einsatz des Etymologiefelds mitsamt XML-Markup bedingt zudem a) den Erhalt der Funktion bei Einsatz des Feldes oder b) das Löschen der Funktion aus dem System.

3.3.3. Der „Graphienblock“

In diesem Schritt wird der Graphienblock des Artikels festgelegt, der alle Graphien des Lemmas umfasst.

Hierzu werden vom System alle Primärzettel des Lemmas innerhalb der zu einem Lemma gehörenden Zettel selektiert. Die selektierten Zettel werden dann in Form einer Tabelle und anhand einer ausgewählten Sortierreihenfolge angezeigt:

Bsp. graphische Varianten
comensament,
commensament zum
Lemma COMENCEMENT.

Alphabetisch: alphabetische Sortierung der Zettelwörter, innerhalb des Zettelwortes
Sortierung nach dem Datum des Sigels jeden Zettels.

Sigel: Sortierung nach Sigel, innerhalb des gleichen Sigels Sortierung alphabetisch
nach dem Zettelwort.

3. Funktionale Anforderungen

Manuell: Zusätzlich kann manuell eine Sortierreihenfolge erstellt werden, die dann bei der Anzeige vorrangig verwendet wird.

Die Erstellung des Graphienblocks kann, wenn nicht manuell nachgearbeitet wird, zu 100% automatisiert erfolgen.

Unterschiede zu den Funktionen des DEAF-Redaktionssystems

Hier gilt es, wie im Wörterbuchblock, die Tabellen zur Zettelanzeige anzupassen (s. Abschnitt 3.3.2.2). Möglicherweise kann der Graphienblock darüberhinaus vollständig vom DEAF-Redaktionssystem übernommen werden.

3.3.4. Der „Bedeutungsblock“

Der Bedeutungsblock des Kurzaufsatzes strukturiert das (polyseme) Artikelwort in seine einzelnen Bedeutungen. Der Bedeutungsblock besteht aus (beliebig vielen, aber mindestens einer) Hauptbedeutung, die jeweils in entweder keine, eine, oder mehrere Unterbedeutungen aufgeteilt ist. Dabei ist die Anzahl der Ebenen der Untergliederung nicht festgelegt; sie kann über eine beliebig tiefe Ebene-Unterebene-Struktur modelliert werden.

Auf XML-Seite wird dabei die oberste Ebene mit dem Tag `<sens>` („Bedeutung“) ausgezeichnet, alle weiteren mit `<sous-sens>` („Unterbedeutung“), wobei die XML-Struktur die Schachtelung der Bedeutungen widerspiegelt.

3.3.4.1. Festlegung der Bedeutungsstruktur

Während der Bearbeitung der Bedeutungsstruktur kann jederzeit:

1. eine neue Hauptbedeutung hinzugefügt werden
2. zu einer Hauptbedeutung eine neue Unterbedeutung hinzugefügt werden
3. Unterbedeutung(en) unter eine andere Hauptbedeutung verschoben werden
4. die Reihenfolge der Hauptbedeutungen geändert werden
5. die Reihenfolge der Unterbedeutungen geändert werden
6. eine Unterbedeutung zu einer Hauptbedeutung gemacht werden (dazu wird die „Wurzel“ des Bedeutungsbaums angezeigt und die Unterbedeutung auf diese „gezogen“)
7. eine Hauptbedeutung gelöscht werden, sofern mehr als eine vorhanden ist (dies ist nur möglich, wenn der Hauptbedeutung keine Unterbedeutungen zugeordnet sind)
8. eine Unterbedeutung gelöscht werden
9. Zettel hinzugefügt/entfernt/verschoben werden

Das Freitextfeld zur Eingabe der Bedeutungen ermöglicht die Eingabe von semantischem Markup wie in Abschnitt 3.3.6 beschrieben. Die eingegebene `definition` wird vom Nutzer manuell gesetzt. In diesem Textfeld erfolgt auch die Festlegung der onomasiologischen Kategorie.

3.3.4.2. Zuordnen der Zettel zu Bedeutungen

Falls die Unterscheidung in Primär-, Sekundär- und Tertiärzettel des DEAF übernommen wird, werden die Funktionen des Zettelzuordnens denen des DEAF-Redaktionssystems entsprechen, d.h. wie folgt:

Zur Unterstützung bei der Ermittlung der Bedeutungsstruktur werden vom Redaktionssystem drei Listen angezeigt, aus denen die Zettel zur Zuordnung in die Bedeutungsstruktur „gezogen“ werden können; diese enthalten jeweils die noch nicht zugeordneten Zettel zur Primär-, Sekundär- und Tertiärliteratur. Welcher Inhalt der Zettel in der jeweiligen Zetteltabelle auf der linken und auf der rechten Seite der Maske angezeigt wird, muss den Zetteln des DAG angepasst werden. Geplant ist es, zu jedem Zettel den Inhalt der Tags `entry`, `pos`, `siglum`, `reference`, `definition` und evtl. `enc-def`, `ref-onomas` und `onomasiology`, in dieser Reihenfolge, anzuzeigen. Die Anzahl der relevanten Informationen für die Erstellung eines semantischen Baumes und für die Zettelzuordnung ist größer als die der DEAF-Zettel, was am onomasiologischen Aspekt des Wörterbuchs liegt. Das bedeutet, dass vergleichsweise breite Bildschirme eingesetzt werden sollten, um eine benutzerfreundliche Bearbeitung des Bedeutungsblockes zu erreichen.

Sobald ein Zettel manuell aus einer dieser Listen ausgewählt und in die Bedeutungsstruktur einsortiert wurde, wird er aus der Tabelle (auf der linken Seite der Maske) entfernt und in der Tabelle auf der rechten Seite angezeigt. Wenn der Zettel aus der Bedeutungsstruktur gelöscht wird, erscheint er automatisch wieder in der entsprechenden Tabelle auf der linken Seite.

Für die Zuordnung der Zettel kann der Redaktor drei Funktionen verwenden.

1. Mithilfe einer *drag and drop*-Funktion kann er einzelne Zettel mit dem Mauszeiger „greifen“ und auf die entsprechende Bedeutung im Bedeutungsbaum „ziehen“.
2. Wenn der Redaktor eine Bedeutung des Bedeutungsbaums aktiviert (i.e. anklickt), erscheint am Ende jeder Tabellenzeile auf der linken Maskenseite ein Kästchen. Durch das Anklicken dieser Kästchen lassen sich alle Zettel markieren, die dann gleichzeitig mit einem Klick der aktiven Bedeutung zugeordnet werden können.
3. Das „Ziehen“ des Schildchens „Alle Zettel“ ermöglicht das gleichzeitige Zuordnen aller auf der linken Seite verbleibenden Zettel zu einer Bedeutung.

Auch beim manuellen Zurücksortieren der Zettel aus der Bedeutungsstruktur in die linke Tabelle können mehrere Zettel zugleich erfasst werden.

Die Sortierung innerhalb der Tabellen erfolgt:

- chronologisch für die Primärliteratur,
- mit einer festgelegten Ordnung (oder alphabetisch) für die Sekundärliteratur,
- mit einer festgelegten Ordnung für die Tertiärliteratur.

Zusätzlich kann manuell die Reihenfolge der Zettel verändert werden. Jeder Haupt- und jeder Unterbedeutung kann ein oder können mehrere Zettel zugeordnet werden. Das Redaktionssystem erlaubt dabei auch, einer Bedeutung keinen Zettel zuzuordnen.

Am Ende der Bearbeitung können nicht-zugeordnete Zettel verbleiben, d.h., es ist für eine Publikation des Kurzartikels technisch nicht notwendig, dass alle Zettel der Zettelstrecke auch einer Bedeutung zugeordnet sind.

3. Funktionale Anforderungen

Unterschiede zu den Funktionen des DEAF-Redaktionssystems

Auch hier gilt es, wie im Wörterbuchblock, die Tabellen zur Zettelanzeige anzupassen. Die Anpassung des Bedeutungsblocks hängt, ebenso wie die des Wörterbuchblocks, zudem von der Entscheidung ab, ob die Unterscheidung in Primär-, Sekundär- und Tertiärzettel des DEAF übernommen werden kann (s. Abschnitt 3.3.2.2).

Die größte Anpassung wird dem onomasiologischen Aspekt des DAG geschuldet werden müssen. Das Konzept des DAGpré ist semasiologisch. Es stellt nicht die Wörter zu den einzelnen onomasiologischen Kategorien zusammen wie das DAG (d.h. ein Lemma kann, wenn es mehrere Bedeutungen hat, unter verschiedenen onomasiologischen Kategorien im DAG behandelt und publiziert werden). Der DAGpré sammelt dagegen alle Bedeutungen zu einem Wort (die der DAG den verschiedenen onomasiologischen Kategorien zuordnen würde) unter einem Lemma und stellt sie dort dar. Insofern ist das Konzept vergleichbar mit dem des DEAF. Dennoch muss dem onomasiologischen Aspekt Bedeutung zugemessen werden; es muss eine Möglichkeit geschaffen werden, jede Bedeutung (jedes Wortes / Lemmas) eindeutig mit einer onomasiologischen Kategorie zu verknüpfen.

Dies ist im Hinblick auf die geplanten Suchfunktionen der elektronischen Publikation wichtig: Die Zuordnung der Bedeutungen (die als Definitionen angelegt werden) zu onomasiologischen Kategorien wird das Abfragen des gesamten DAGpré nach Hallig-Wartburg-Kategorien ermöglichen und so den DAGpré nicht nur semasiologisch sondern auch onomasiologisch ausrichten können wie den DAG.

3.3.5. Der „Kommentarblock“

Zu jedem Kurzartikel kann ein Kommentarblock hinzugefügt werden, der in einem Freitext mit Markup-Unterstützung weitere Informationen enthält, i.e. Kommentare jeglicher Art, Hyperlinks zu anderen Artikeln des DAGpré, etc. Das Freitextfeld ermöglicht die Eingabe von semantischem Markup wie in Abschnitt 3.3.6 beschrieben.

In der Konzeption ist der Kommentarblock schlicht, jedoch ermöglicht er dem Redaktor der Kurzartikel das Dokumentieren von wissenschaftlicher Arbeit, das Festhalten von Überlegungen, das Hinweisen auf wichtige Zusatzinformationen, etc.

3.3.6. Semantisches Markup in Freitextfeldern

Im Redaktionssystem gibt es mehrere Textfelder, bei denen der Redaktor semantisches XML-Markup eingeben kann. Dabei unterstützt die Nutzeroberfläche den Redaktor in folgender Weise:

- Eine Liste von XML-Elementen (Tags), die bei Auswahl direkt in den Text eingefügt werden, steht zur Verfügung (s. Anhang D).
- Bei Auswahl eines Tags, für den Attribute angegeben werden können oder müssen, erscheint ein Pop-up-Fenster, in dem die Attribute vorgegeben sind und vom Nutzer ausgefüllt werden.

Die zur Verfügung stehenden Tags sind dabei für alle im System auftretenden Textfelder die selben.

Beim Speichern des eingegebenen Textes prüft ein XML-Parser, ob das eingegebene Markup dem XML-Schema entspricht, sofern dieses für das jeweilige Feld komplett vorliegt.

3.4. Datenkodierung

Sämtliche Texte werden in Unicode abgespeichert. Unter den Varianten UTF-8, UTF-16 und UTF-32 wird dabei diejenige verwendet, die alle verwendeten Zeichen (mit Akzenten etc.) enthält. Für das System des DEAF ist UTF-8 ausreichend; vermutlich wird das auch für das des DAG der Fall sein.

3.5. Verknüpfung mit externen Datenbeständen

Zwei Möglichkeiten der Verknüpfung mit anderen Projekten sind zum Zeitpunkt der Konzepterstellung angedacht.

1. Im Kommentarblock können die Einträge mit externen Lexikographie-Projekten in Form „externer Links“ verknüpft werden, insbesondere mit
 - a) dem *Dictionnaire de l'occitan médiéval* – DOM, Bayerische Akademie der Wissenschaften, s. <http://www.dom.badw-muenchen.de/>,
 - b) der entstehenden digitalen Version des *Französischen Etymologischen Wörterbuchs* – FEW, s. bisher
<https://apps.atilf.fr/lecteurFEW/index.php?r=page/view>
(es ist noch nicht möglich, auf einzelne Artikel zu verlinken; eine vollständig digitale Version ist in Arbeit),
 - c) zu prüfen ist das Material in *Lexic occitan medieval*, publiziert auf
http://www.jfbrun.eu/lengadoc/lexic_medieval.htm,
 - d) zu prüfen sind weiterhin die Projekte, die das Projekt *Lexilogos* unter der Rubrik „ancien occitan“ listet, s.
http://www.lexilogos.com/occitan_ancien.htm.

Momentan wird hierzu die manuelle Eingabe einer URL vorgesehen.

2. Die zweite Möglichkeit ist die der Verknüpfung mit anderen Projekten auf Daten-, i.e. Datenbankebene. Vorbild ist hier das Modell der Verknüpfung des DEAF mit dem altfranzösischen Korpusprojekt *Les plus anciens documents linguistiques de la France* – DocLing. Es muss überlegt werden, welche Wörterbuchprojekte oder Korpusprojekte sich für eine Verschmelzung auf Datenebene eignen. Dies ist nicht Teil des vorliegenden Konzeptes.

3.6. Elektronische Publikation

Das Gesamtziel des Projektes ist es, den gesamten Wortschatz des Altgaskognischen online zur Verfügung zu stellen,

- entweder als DAG in seiner ursprünglichen Form mit zwei Publikationsalternativen
 1. die redigierten Artikel sind online verfügbar als eingescannte Bilddateien (Rastergrafiken), die Lemmata der Artikel sind in eine Suchfunktion integriert.
 2. die redigierten Artikel sind retrodigitalisiert, d.h. vollständig in die Datenbank und in die Onlinepublikation des DAG integriert.

Die Rechte der Onlinepublikation liegen beim DAG.

3. Funktionale Anforderungen

- oder als DAGpré (mit *open access*).

Zusammen mit den Nebenprodukten DAGBiblél und dem digitalen Zettelkasten bilden DAG und DAGpré „DAG électronique“ – DAGél.

Dafür ist es nötig, eine XHTML-Version aus der Datenbank zu exportieren. Dies ist zur Zeit nicht als integrierte Funktion des Redaktionssystems geplant, sondern soll – wie für DEAFél – mit einem externen Script erfolgen. Eine Integration in das System wäre jedoch wünschenswert. Dann könnte auch ein automatisches Umsetzen der innerhalb des Redaktionssystems vorgenommenen Änderungen im XML-Markup erfolgen, s. Abschnitt **3.1.1**.

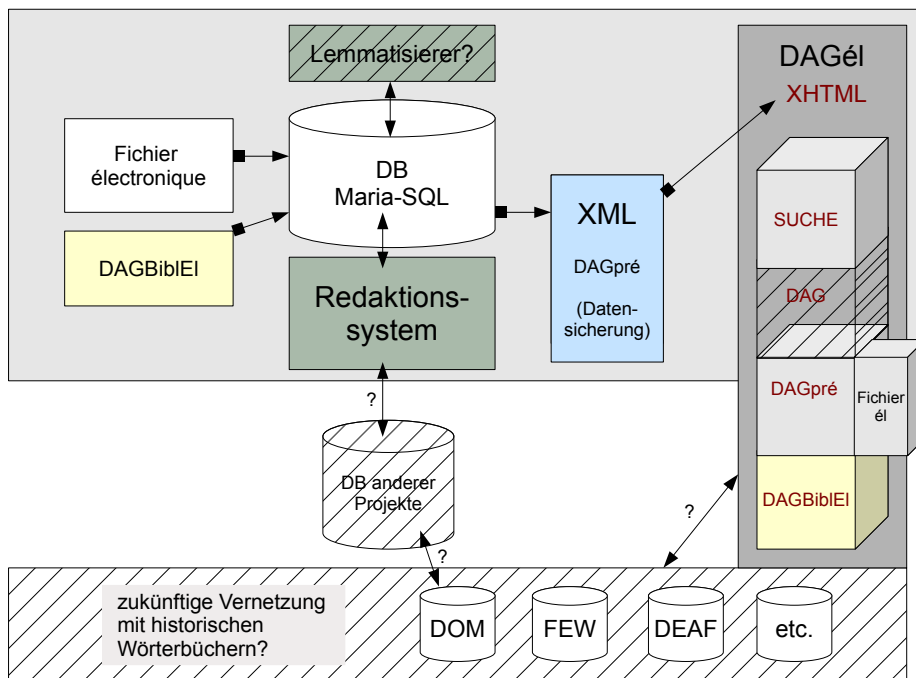
3.6.1. Suchfunktionen

Suchfunktionen sollen vielfältige Zugangsmöglichkeiten zu den Informationen von DAGél bieten:

1. einen alphabetischen Zugang durch das Abfragen von Lemmata,
z.B. *comencement*
 2. einen onomasiologischen Zugang durch das Abfragen von Hallig-Wartburg-Kategorien bzw. Hallig-Wartburg-Begriffen ,
z.B. „B III b ó“ bzw. *début*
 3. einen semantischen Zugang durch das Abfragen von Wörtern aus Definitionen (definition; enc-def).
z.B. *début, entrée (du mois, de l'année)*
z.B. *acomplir lor voluntad de*
z.B. *en prumer comensament*
- Weitere Abfragen, z.B. nach POS, „loc. verb.“, „loc. adv.“, fachsprachliche Auszeichnungen, etc., werden durch die jeweilige XML-Auszeichnung der Daten ermöglicht und sind noch zu spezifizieren.



Überblick Systemarchitektur



B

Datumsangaben

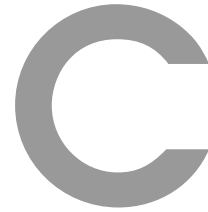
Es gibt im Altgaskognischen viele Texte, die sich z.B. anhand einer Datumsangabe im Text selbst exakt auf ein Entstehungsjahr datieren lassen, z.B. **1210** oder **1389**.

Andere altgaskognische Texte lassen sich nur annäherungsweise auf einen bestimmten Entstehungszeitraum datieren, z.B. auf das 2. Drittel des 13. Jahrhunderts. Diese Zeitraum-Datierungen müssen in eine exakte Jahreszahl umgerechnet werden, und zwar auf das arithmetische Mittel (aufgerundet) des jeweiligen Zeitraums. Diese Umrechnung ist nötig, damit das System in der Lage ist, die zu den Texten gehörenden Zettel in der korrekten chronologischen Reihenfolge zu sortieren und in den Tabellen darzustellen (s. „Graphienblock“, 3.3.3, „Bedeutungsblock“, 3.3.4).

Tabelle B listet die Umsetzung am Beispiel des 13. Jahrhunderts auf.

Zeitraum-Datierung	Errechnete Jahreszahl
déb. 13 ^e s.	1210
1 ^{er} q. 13 ^e s.	1213
1 ^{er} t. 13 ^e s.	1217
1 ^e m. 13 ^e s.	1225
2 ^e q. 13 ^e s.	1238
mil. 13 ^e s.	1250
2 ^e t. 13 ^e s.	1250
13 ^e s.	1250
3 ^e q. 13 ^e s.	1263
2 ^e m. 13 ^e s.	1275
3 ^e t. 13 ^e s.	1284
4 ^e q. 13 ^e s.	1288
fin 13 ^e s.	1290

Tabelle B.1.: Liste der Umsetzung von Zeitraum-Datierungen in Jahreszahlen, am Beispiel des 13. Jhs.



Part-of-Speech Tags

Die nachfolgende Liste zeigt die verwendeten Wortklassen (POS-Tags) des DEAF. Evtl. kann sie in der Form vom DAGpré-Projekt übernommen werden.

f. m. s. s.f. s.m. m.pl. f.pl. adj. adj.poss. adj. substantivé adv. art.déf. art.indéf. prép. interj. conj.	v. v.tr. v.intr. v.pron. v.pron. à valeur neutre v.pron. à valeur réfléchie v.pron. à valeur réciproque v.pron. à valeur passive v.tr. + c.o.i. v.tr. + c.o.d. emploi abs. inf. substantivé p.p. p.p. comme adj. p.p. pris comme adj. p.prés. p.prés. pris comme adj.	pron. pron.pers. pron.dém. pron. indéfini pron.interrog. pron.poss. pron.rel. pron.adj. loc. loc. adj. loc. adv. loc. verb. loc. subst. loc. interj. loc. prép. loc. pron. ?
--	---	--



XML-Elementeliste

Die nachfolgende Liste zeigt die verwendeten XML-Elemente für das semantische Markup innerhalb der Freitextfelder des DEAF. Evtl. kann sie in der Form vom DAGpré-Projekt übernommen werden.

a br case case-ending case-form citation cited-letters cited-word collocation comment compound
definition designation evidence flexion flexion-ending footnote gram idem link locution manuscript
onomasiology pos proverb quotation remark remark-reference sc siglum sup terminology
text-reference usage



Hauptmenü

Übersicht

Zettelkästen

Sigel

Onomasiologische Kategorien

Benutzerverwaltung

Abmelden

Übersicht

Datenpflege

[Zettelkästen](#)

[Sigel](#)

[Kategorien](#)

Verwaltung

[Benutzerverwaltung](#)



Zetteleingabemaske

Zettel bearbeiten

Sigel	<input type="text" value="FEW"/>	Datierung	<input type="text"/>	Einzeldatum	<input type="text"/>
Stelle	<input type="text"/>	Lokalisierung	<input type="text"/>	Sprache	<input type="text"/>
Wort	<input type="text" value="bla"/>	Definition	<input type="text"/>		
Wortart	<input type="text"/>	Kontext	<input type="text"/>		
Lemma	<input type="text"/>				
Kategorie	<input type="text"/>				
Begriff	<input type="text"/>				



Eingabemaske der Bibliographie

Sigel bearbeiten

[Sigelvorgänger](#) [Sigelvarianten](#)

Name

Datierung

Lokalisierung

Sprache

Text

Eingabe nicht erforderlich