













Protokoll für die Codierung der Datenbankinhalte

Änderungsprotokoll	2
Voreinstellungen in Oxygen:	2
Zur Arbeitsweise:	2
Verknüpfen bzw. Gruppieren von Textelementen zu einem Datensatz in xml	3
Button-Menüleiste:	3
a) GeoObjekt-Texte	3
b) Personen, Orte, Berufe	4
c) Sachbegriffe.....	4
Bearbeitung der GeoObjekte	5
a)  Kartensymbol der Reinschrift eines GeoObjekts:	5
b)  Beschreibungstext eines GeoObjekts	8
c)  Arealgröße eines GeoObjekts.....	8
d)  Kartensymbol der Urschrift eines GeoObjekts.....	8
e)  Kategorien der GeoObjekte.....	9
Bearbeitung der Sachbegriffe	9
a) Vorbereitung.....	9
b) Kategorien	10
c) Normalisierung der Sachbegriffe.....	11
Bearbeitung der Personen	11
a)  ,  ,  Nachname einer Person	12
b)  Vorname einer Person	12
c)  Titel einer Person.....	13
d)  Beruf einer Person	13
Bearbeitung der Orte	13
a)  Ort in Schwedisch Pommern	13
b)  Ort außerhalb Schwedisch Pommerns	14

Änderungsprotokoll

Montag, 3. Dezember 2009

- neue Areal-Kategorie für die GeoObjekte: "Hofstellen (bewohnt u. wüst) +/- Koppeln"
- Bearbeitung der Sachbegriffe

Freitag, 8. Januar 2010

- Bearbeitung der Personen
- Bearbeitung der Orte

Das Protokoll spiegelt den aktuellen Arbeitsstand wieder und wird bei Bedarf weiterentwickelt. Sollten z.B. in den Texten Phänomene auftreten, die mit der bisherigen Methode nicht erfasst werden können oder gibt es Unsicherheiten bzgl. der Auszeichnungen, dann bitte Rücksprache halten.

Voreinstellungen in Oxygen:

- 1.) Den Ordner "Oxygen_AutorTools" unter C:\ kopieren
- 2.) Programm "Oxygen XML-Editor" öffnen
- 3.) Optionen -> Importiere globale Optionen wählen und im Fenster Öffnen die Datei "C:\C_Oxygen_AutorTools\globaleOptionen.xml" öffnen
- 4.) erfolgreichen Import bestätigen

Zur Arbeitsweise:

- die zu bearbeitende xml-Datei im Oxygen XML Editor öffnen
- von Textmodus in den Autorenmodus wechseln

- bevor mit der Arbeit an einer XML-Datei begonnen wird, eine Kopie der Datei in den Ordner "alte Dateien" ablegen, um den letzten Arbeitsstand zu sichern.
- der Arbeitsstatus wird in der Datei "Dok_Bearbeitungsstand.xls" (P:\08 Schwedische Landesaufnahme) festgehalten
- der Arbeitsordner ist P:\08 Schwedische Landesaufnahme\08WebTexte\xhtml-Edition. Darin liegen alle aktuellen Dateien, die in gewissen Zeitabständen auf die Homepage überspielt werden. Die sich darin befindlichen Dateien bitte nur innerhalb dieses Ordners und auf P bearbeiten.

- Oxygen wurde so an die Arbeiten für das DFG-Projekt angepasst, dass die Markierungen zu 95 % per Button-Menüleiste im Autorenmodus vorgenommen werden können. Die **Autorenansicht ist keine echte WYSIWYG Umgebung**, sondern nur eine optische Hilfe im Umgang mit xml-Dateien. Alle Tags aus der vorangegangenen editorischen Bearbeitung der Texte sind in der Datei enthalten, werden aber ausgeblendet und sind zum Teil nur sichtbar

durch die optische Unterteilung des Textes im Autorenmodus (wie Überschriften, Kapitel, Kapitelüberschriften, Paragraphen, Zeilenumbrüche u.s.w.).

- Außer einem korrigierenden Eingriff sind Veränderungen an der Textstruktur unzulässig.
- Die xml-Dateien werden von den Editoren wohlgeformt und valide übergeben, mit einer Ausnahme die Kartensymbole der Urschrift in den Fußnoten betreffend ([näheres siehe unten](#)).

Nach der Bearbeitung der Datenbankinhalte in xml muss die Datei ebenfalls fehlerfrei, also valide und wohlgeformt sein.

- Für die Fehlersuche und -breinigung ist die Autorenansicht ungeeignet, weil sie nur eine eingeschränkte Sicht auf die xml-Struktur zulässt, so dass für Korrekturarbeiten der Textmodus empfohlen wird.

- Im Autorenmodus kann das Taggen im Text mit Hilfe der projektbezogenen Buttons in der Menüleiste erfolgen. Dabei wird die entsprechende Textstelle markiert und nach dem Mausklick auf einen Button, wird die markierte Textstelle mit einem TEI-Tag umgeben. Bei einigen Elementen ist noch das nachträgliche Bearbeiten von Attributwerten notwendig, vor allem bei Daten, die zueinander in Beziehung zu setzen sind (z.B. Nachname und Vorname einer Person). Die Attribute können ebenfalls über die Ansicht "Attribute" im Autorenmodus bearbeitet werden.

- Generell sollte man bei der Bearbeitung der Texte gedanklich und technisch so arbeiten, dass zum einen das Ergebnis am Ende TEI und XML konform ist aber zugleich auch das aus den XML-Dateien komplett automatisch fehlerfreie Datensätze für die Datenbank generiert und überspielt werden können. Eine manuelle Nachbereitung der Datenbank für etwaige Spezialfälle ist unzulässig.

- **Beim Taggen der Texte gilt, das prinzipiell nur vollständige Wörter oder Wortgruppen markiert werden** (inklusive beispielsweise von Deklinations- oder Pluralendungen).

Überflüssige Satzzeichen am Ende eines Wortes oder einer Wortgruppe werden nicht markiert.

Bsp. für **falsche** Markierung: "`<rs type="profession">Kossat</rs>en.`"

Bsp. für **richtige** Markierung: "`<rs type="profession">Kossaten</rs>.`"

Das Auszeichnen von Begriffen, die zu mehreren Kategorien zugeordnet werden müssen, erfolgt durch Verschachteln der Tags.

Bsp. für **falsche** Markierung:

`"<rs type="plant">Erlen</rs><rs type="miscellanea">bruch</rs>"`

Bsp. für **richtige** Markierung:

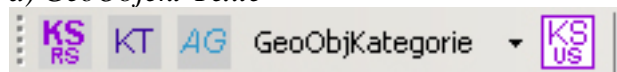
`"<rs type="miscellanea"><rs type="plant">Erlenbruch</rs></rs>"`




Verknüpfen bzw. Gruppieren von Textelementen zu einem Datensatz in xml

Einige markierte Textstellen stellen bei der Weiterverarbeitung der Daten eine Informationseinheit dar und müssen als solche in xml gruppiert werden. Betroffen davon sind die markierten Texte bzgl. der GeoObjekte und Personen. Die Verknüpfung wird in xml über Attribute realisiert. Den Ankerpunkt bildet dabei das Attribut „xml:id“ und die Verknüpfung zum Ankerpunkt wird über das Attribut „ref“ angegeben. Der Wert von „xml:id“ muss dabei für das gesamte Dokument eindeutig sein. Näheres siehe [Bearbeitung der GeoObjekte](#) und [Personen](#).

Button-Menüleiste:

a) GeoObjekt-Texte



-  = Kartensymbol der Reinschrift eines GeoObjekts
-  = Beschreibungstext eines GeoObjekts
-  = Arealgröße eines GeoObjekts





Kategorien der GeoObjekte:


GeoObjKategorie ->	
	Acker
	wüster Acker
	Wiesen
	Wald und Weiden
	Obst-, Gemüse- und Kräutergärten
	Hofstellen (bewohnt) u/o. Wegstücke +/- Gärtchen
	Sonstiges
	Gebäude
	Gewässer
	Wald
	Weide
	Hofstellen (wüst/unbewohnt)
	Hofstellen (bewohnt u. wüst) +/- Koppeln


-  = Kartensymbol der Urschrift eines GeoObjekts

b) Personen, Orte, Berufe



-  = Vorname/n einer Person
-  = Nachname einer Person (ortsansässig)
-  = Nachname einer Person (ortsfremd)
-  = Titel einer Person

-  = Wdhl. des Nachnamens einer Person

-  = Beruf (evtl. einer Person)

-  = Ort in Schwedisch Pommern

-  = Ort außerhalb Schwedisch Pommerns

c) Sachbegriffe



-  = Abgaben, Steuern und Dienste

-  = Handdienste


-  = Spanndienste

 = Hand-und Spanndienste

 = Grundbesitzverhältnis

 = Landnutzung

 = Betriebsform, Gebäude und Monumente

 = Tiere

 = Pflanzen

 = Landschaftliche Merkmale, Sonstiges


Bearbeitung der GeoObjekte

Das Ausweisen der Textabschnitte bzgl. der GeoObjekte ist das inhaltlich strukturierte Erfassen der Textabschnitte, die zu Arealen auf den Karten korrespondieren. Zum überwiegenden Teil betrifft es die Texte aus den Dokumentbereichen "Arealsausrechnung" und "Revision" aber auch die nummerierte Aufzählung der Einwohner in der allg. Ortsbeschreibung haben Entsprechungen zu Karteneinträgen.

Bis zu 5 verschiedene Elemente finden in diesem Bereich eine Anwendung. Neben dem Markieren der Texte müssen die einzelnen Elemente auch noch zueinander in Beziehung gesetzt werden um aus den Einzelementen zusammengehörende Datensätze für die Datenbank generieren zu können. Die Dateninhalte im Bereich der GeoObjekte bilden die Grundlage für das WebGIS. Besondere Aufmerksamkeit sollte unbedingt geübt werden bei der Bearbeitung der GeoObjekt-Texte, die bereits manuell in die Datenbank eingearbeitet und im WebGIS eingebunden wurden. (Das betrifft alle unsere alten doc-Dateien, die in xml umgearbeitet werden müssen.) Denn hier dürfen keine Veränderungen der Datenbankinhalte bzgl. ihrer Zuordnung stattfinden, mit Ausnahme von Korrekturarbeiten. Ein GeoObjekt ist einer Kategorie untergeordnet besteht idealerweise aus einem "Kartensymbol der Reinschrift", dem "Beschreibungstext", der "Arealgröße" und evtl. kann noch ein von der Reinschrift verschiedenes "Kartensymbol der Urschrift" erwähnt sein.

Bzgl. der Anordnung der GeoObjekte im xml-Text: siehe Protokoll_xml_Edition.doc (P:\08 Schwedische Landesaufnahme\08WebTexte\Protokolle für xml\Protokoll_Übers Editoren)

a)  *Kartensymbol der Reinschrift eines GeoObjekts:*

Das Kartensymbol der Reinschrift wird bei Verwendung des Buttons  mit folgendem Tag umgeben:

```
<rs type="mapSymbol_fairCopy" xml:id="msymb_n/o[n]">[Kartensymbol]</rs>
```

Wie bereits oben erwähnt, benötigen wir für die GeoObjekte auch ein Attribut, welches eine ID enthält, so dass jedes GeoObjekt eindeutig ausgewiesen wird. So können verschiedene Textbestandteile eines GeoObjektes einander zugeordnet werden, unabhängig von ihrer Position im Text. Das Kartensymbol der Reinschrift hat dabei eine zentrale Bedeutung, denn es enthält das Attribut xml:id, in dem die ID des GeoObjektes zugewiesen wird. Für jedes GeoObjekt muss immer eine xml:id vergeben werden, also auch für die seltenen Textstellen eines GeoObjektes ohne ein im Beschreibungstext erwähntes Kartensymbol der Reinschrift. In diesen Fällen wird ein leeres Tag mit der ID eingefügt (Hinweis: im Textmodus arbeiten, da leere Objekte im Autorenmodus schnell übersehen werden können):

```
<rs type="mapSymbol_fairCopy" xml:id="msymb_n/o[n]"/>
```

Der voreingestellte Wert „msymb_n/o[n]“ des Attributes „xml:id“ dient nur als Platzhalter und muss durch einen eindeutigen Wert ersetzt werden. (Wird er nicht verändert, erzeugt er beim Validieren der XML-Datei eine Fehlermeldung.) Bezüglich der Eindeutigkeit der Werte ist anzumerken, dass die ID nicht allein durch die Nummern gebildet werden, sondern in Kombination mit einer Zeichenkette. So ist die ID "msymb_n1" verschieden von "msymb_o1".

Kartensymbole können Zeichenketten wie "Aa" oder "A1" sein, aber auch Symbole, die in XML wie folgt kodiert werden: `<g ref="#symbol_27"/>` z.B. für das Symbol 27. Bei der Auszeichnung der GeoObjekte müsste das Symbol mit dem entsprechenden Tag für das Kartensymbol umschlossen werden:

```
<rs type="mapSymbol_fairCopy" xml:id="msymb_n/o[n]"><g ref="#symbol_27"/></rs>
```

Zwei Arbeitsweisen sind dabei zu unterscheiden:

1.) **Vergabe von ID-Werten für die alten Übersetzungstexte**, die bereits manuell in die Datenbank eingetragen wurden:

Die Vergabe der ID für die Textinhalte der GeoObjekte, die bereits in der Datenbank erfasst sind, orientiert sich an der ID, die den Texten in der Datenbank zugewiesen wurden. Weil diese Texte bereits in der WebGIS-Anwendung gefunden werden und eine feste Position in der Datenbank haben, ist es wichtig, diese Positionsangabe unverändert in die XML-Datei zu übernehmen, damit beim automatisierten Update der Daten, die Datenbankinhalte nicht vertauscht werden und damit das WebGIS nachträglich falsche Daten übermittelt bekommt. Deshalb diesen Bereich bitte äußerst aufmerksam und mit großer Sorgfalt bearbeiten!

Bzgl. der Arbeitsweise empfiehlt es sich, die Datenbank-Ansicht der GeoObjekte und den XML-Editor untereinander zu legen. Über den Orts-Filter werden die GeoObjekte des zu bearbeitenden Ortes ausgewählt. Dann wird die ID des GeoObjektes in der Datenbank herausgesucht, die mit den Textstellen in der XML-Datei übereinstimmt, und das Reinschrift-Kartensymbol bekommt die ID mit vorangestelltem "msymb_o" zugewiesen.

- Bsp. für "Acker A1 von Spiegelsdorf": (die ID in der Datenbank lautet für dieses GeoObjekt: 1142)

```
<rs type="mapSymbol_fairCopy" xml:id="msymb_o1142">A1</rs>
```

Wurde ein GeoObjekt in XML mit der Datenbank-ID versehen, dann bekommt der Datensatz in der Datenbank einen Eintrag in "xml_Status":

"ok" wird eingetragen, wenn die in der XML-Datei markierten Textstellen eines GeoObjektes mit den Einträgen in der Datenbank übereinstimmen

"geändert" wird eingetragen, wenn durch Korrekturen inhaltliche Abweichungen in der XML-Datei bzgl. der markierten Textstellen eines GeoObjektes bestehen im Vergleich zu den Einträgen in der Datenbank, so dass es nach der (automatisierten) Überspielung der Daten aus den XML-Dateien zu geänderten Einträgen in der Datenbank kommt. (Die Übersetzungen der XML-Dateien stellen die aktuellste Version eines Textes dar, so dass die Datenbankinhalte durchaus veraltet sein können.) Wichtig: Sollten durch die Arbeit mit den XML-Dateien Korrekturen notwendig sein, so können diese nur inhaltlich vorgenommen werden, denn der ursprüngliche Bezug zum

GeoObjekt darf nicht verändert werden. (Aus dem obigen Beispiel müsste der Bezug zur Fläche A1 von Spiegelsdorf und der ID 1142 unbedingt erhalten bleiben.)

"gelöscht" wird eingetragen, wenn durch Korrekturen ein Datensatz in der Datenbank keine Entsprechung in xml findet.

Bsp. (Hofstellen in Spiegelsdorf):

G Am Hof des Verwalters ist ein Garten, der nun aber sehr verwildert ist. Die Bauern besitzen Obstgärten mit Apfel-, Birn-, Kirsch- und Pflaumenbäumen außerdem auch einige Hopfenstangen. Auch der Hirte hat bei seiner Hütte einen Obstgarten, der zusammen mit eines jeden Gemüsegarten und der Hofstelle umfasst 3M 90R
Die Straße ist 1M 180R ~~4M 270R~~

In der Datenbank existieren für das obige Bsp. zwei Einträge (ID 1162 [Hof und Gärten] und 1163 [Straße]). Da eine Gesamtsumme zu G existiert und die Straße kein eigenes Symbol hat, gehören beide Texte zu G, also zu einer einzigen ID. So bekommt G in der xml-Datei die ID 1162 für die Fläche G aus der Datenbank und der Beschreibungstext umfasst den ganzen blau eingefärbten Text einschließlich der Teil-Arealgrößen. Als Arealgröße bekommt G die Gesamtgröße zugewiesen. In der Datenbank wird in "xml_Status" beim Datensatz mit der ID 1162 "geändert" und beim Datensatz mit der ID 1163 "gelöscht" eingetragen, da letzterer durch die Korrekturen keine Entsprechung zur xml-Datei mehr hat.

Sollten in der xml-Datei GeoObjekte kodiert sein, die in der Datenbank noch nicht erfasst wurden, muss eine **neue ID** erstellt werden, indem neue ID's mit vorangestelltem "msymb_n" aufsteigend von 1 bis n durchnummeriert werden. Die ID's müssen innerhalb einer Ortsbeschreibung eindeutig sein.

- Bsp. für "die Wiese D[11-12] von Spiegelsdorf": (neue ID 16 vergeben, weil das GeoObjekt bisher nicht in die Datenbank aufgenommen wurde)

```
<rs type="mapSymbol_fairCopy" xml:id="msymb_n16">D[11-12]</rs>
```

2.) **Vergabe von ID-Werten bei neuen Übersetzungstexten**, die noch nicht in der Datenbank erfasst sind:

Für die GeoObjekt-Texte, die noch nicht in der Datenbank erfasst wurden, muss eine ID erstellt werden, indem aufsteigend von 1 bis n durchnummeriert und davor die Zeichenkette "msymb_n" geschrieben wird. Die ID's müssen innerhalb einer Beschreibung eindeutig sein.

- Bsp.: (neue ID's 1 bis 5 vergeben)

```
<rs type="mapSymbol_fairCopy" xml:id="msymb_n1">A</rs>
```

```
<rs type="mapSymbol_fairCopy" xml:id="msymb_n2">A1</rs>
```


```
<rs type="mapSymbol_fairCopy" xml:id="msymb_n3">A2</rs>
```


```
<rs type="mapSymbol_fairCopy" xml:id="msymb_n4">B</rs>
```

```
<rs type="mapSymbol_fairCopy" xml:id="msymb_n5">C</rs>
```

Zusammenfassend sei noch einmal darauf hingewiesen, dass im Attribut "xml:id" eine ID mit der Zeichenkette "**msymb_o**" für eine ID aus der Datenbank steht und eine ID mit der

Zeichenkette "msymb_n" für eine neue ID, deren GeoObject, das es repräsentiert noch nicht in die Datenbank eingetragen wurde.

b)  *Beschreibungstext eines GeoObjekts*


Der Beschreibungstext eines GeoObjekts wird bei Verwendung des Buttons  mit folgendem Tag umgeben:


```
<rs type="area_text" ref="#msymb_n/o[n]">[ Beschreibungstext]</rs>
```

Der voreingestellte Wert „#msymb_n/o[n]“ des Attributes „ref“ dient nur als Platzhalter und muss durch einen Wert ersetzt werden. Der richtige Wert ist der Wert des Attributs *xml:id* des zum Beschreibungstext gehörenden Kartensymbols der Reinschrift, mit dem vorangestelltem Zeichen "#".

Bsp.: lautet die *xml:id* des Reinschrift-Kartensymbols "msymb_n1", dann ist der richtige Wert für das Attribut ref "#msymb_n1".

Sollten z.B. durch geklammerte Texte ein Beschreibungstext für mehrere GeoObjekte gelten, dann kann ref auch mehrere Referenz-Ids enthalten, die durch ein Leerzeichen voneinander getrennt sind, z.B.: ref="#msymb_n1 #msymb_n2 #msymb_o4567 #msymb_o346".

c)  *Arealgröße eines GeoObjekts*

Die Arealgröße eines GeoObjekts wird bei Verwendung des Buttons  mit folgendem Tag umgeben:


```
<rs type="area_size" ref="#msymb_n/o[n]">[ Arealgröße]</rs>
```

Der voreingestellte Wert „#msymb_n/o[n]“ des Attributes „ref“ dient nur als Platzhalter und muss durch einen Wert ersetzt werden. Der richtige Wert ist der Wert des Attributs *xml:id* des zur Arealgröße gehörenden Kartensymbols der Reinschrift, mit dem vorangestelltem Zeichen "#".

Bsp.: lautet die *xml:id* des Reinschrift-Kartensymbols "msymb_n1", dann ist der richtige Wert für das Attribut ref "#msymb_n1".

Sollten z.B. durch geklammerte Texte eine Arealgröße für mehrere GeoObjekte gelten, dann kann ref auch mehrere Referenz-Ids enthalten, die durch ein Leerzeichen voneinander getrennt sind, z.B.: ref="#msymb_n1 #msymb_n2 #msymb_o4567 #msymb_o346".

d)  *Kartensymbol der Urschrift eines GeoObjekts*

Das Kartensymbol der Urschrift eines GeoObjekts wird bei Verwendung des Buttons  mit folgendem Tag umgeben:

```
<rs type="mapSymbol_originalText" ref="#msymb_n/o[n]">[ Kartensymbol der Urschrift]</rs>
```

Der voreingestellte Wert „#msymb_n/o[n]“ des Attributes „ref“ dient nur als Platzhalter und muss durch einen Wert ersetzt werden. Der richtige Wert ist der Wert des Attributs *xml:id* des zum Urschrift-Kartensymbol gehörenden Kartensymbols der Reinschrift, mit dem vorangestelltem Zeichen "#".

Bsp.: lautet die *xml:id* des Reinschrift-Kartensymbols "msymb_n1", dann ist der richtige Wert für das Attribut ref "#msymb_n1".

Bzgl. des Kartensymbols der Urschrift kann die Besonderheit auftreten, dass es in der Fußnote erwähnt wurde, die aber in der Autorenansicht ausgeblendet und somit auf den ersten Blick nicht ersichtlich ist. Solche Kartensymbole wurden bereits von den Editoren markiert. Am Ende der Markierung der GeoObjekte empfiehlt es sich daher in den Textmodus zu wechseln und nach der Zeichenkette:

"<rs type="mapSymbol_originalText" ref="#msymb_n/o[n]">" zu suchen. Sollten Urschrift-Kartensymbole in den Fußnoten markiert worden sein, dann muss in das Attribut ref der richtige Wert eingetragen werden.

e) **GeoObjKategorie** Kategorien der GeoObjekte

Jedes GeoObjekt wird einer Arealkategorie zugewiesen. Im Unterschied zur bisherigen Vorgehensweise bei der manuellen Dateneingabe, richtet sich die Zuordnung nach den Überschriften, wie sie im Originaltext zugeordnet wurden. Also eine Weide, die sich unter der Überschrift Wiese befindet bekommt die Kategorie Wiese zugeordnet. Nur bei mehrfach Überschriften wie "Wald und Weide" oder bei zusammenfassenden Überschriften wie "Hofstellen", worunter sich auch meistens Wege und Gärten befinden bekommen die GeoObjekte ihre Zuordnung gemäß ihres Textinhaltes.

Folgende Kategorien können momentan ausgezeichnet werden:

GeoObjKategorie ->	Acker wüster Acker Wiesen Wald und Weiden Obst-, Gemüse- und Kräutergärten Hofstellen (bewohnt) u/o. Wegstücke +/- Gärtchen Sonstiges Gebäude (betrifft Personen mit Nummern, o.a.) Gewässer Wald Weide Hofstellen (wüst/unbewohnt) Hofstellen (bewohnt u. wüst) +/- Koppeln
--------------------	--

Sollten weitere Kategorien notwendig sein, dann bitte Rücksprache halten.

Bearbeitung der Sachbegriffe

a) Vorbereitung

Zunächst werden Sachbegriffe nur als solche markiert, wenn sie auch wirklich zur Gemarkung gehören und nicht durch Verneinungen auszuschließen sind.

[Siehe oben](#) zur Arbeitsweise: keine einzelnen Wortbestandteile markieren, sondern nur ganze Wörter oder Wortgruppen. Vom Verständnis her werden in den xml-Texten die originalen Begriffsvorkommen markiert, wie sie im Textumfeld stehen und nach dem Datenbanktransfer werden die Originalbegriffe in der Datenbank normalisierten Begriffen zugeordnet, die dann dem Nutzer in der Suchmaske angeboten werden können.

Bei der Kennzeichnung der Sachbegriffe für die Suchmaschine können die Inhalte der Datenbank genutzt werden. Um den Arbeitsaufwand zu verringern und die Bearbeitung der Dateien zu erleichtern, kann die xml-Datei automatisch mit den Datenbankinhalten abgeglichen und (vor)getaggt werden: näheres siehe dazu im Protokoll "[Protokoll automVortaggen.doc](#)" im gleichen Verzeichnis.

Nach dem Automatischen Taggen des Textes müssen die Markierungen auf Richtigkeit geprüft werden und gleichzeitig Korrekturen und Ergänzungen eingearbeitet werden.

Prinzipiell werden alle Vorkommen eines Sachbegriffs im Text ausgewiesen.

b) Kategorien

Die Sachbegriffe werden folgenden Kategorien zugeordnet:

B Beruf (evtl. einer Person)

Berufe, die sich auf eine konkrete Person beziehen lassen, bekommen als Attribut eine Personenreferenz zugeordnet ([siehe unten](#)). Ohne Personenbezug werden sie einfach als Sachbegriff markiert.

Abg Abgaben, Steuern und Dienste

Markierung von Abgaben und Steuern. Die Dienste werden in 3 Untergruppen geteilt und in sinnvollen Wortgruppen markiert, die die Informationen zu den einzelnen Diensten enthalten:

HD Handdienste

- Dienste mit Personen und ohne Tiere (Mähen, Dreschen, Binden, etc.)

SP Spanndienste

- Dienste mit Tieren und ohne Personen (Pflügen, Fuhrdienst, etc.)

HD SD Hand-und Spanndienste

- Dienste mit Personen und Tieren (Bsp.: Vollbauer dient mit drei Personen, zwei Ochsen und drei Pferden)

Beispiel für die Markierung von "Hand-und Spanndienste":

```
<rs type="duty"><rs type="service_mixed">Jeder Vollbauer dient zur Saat mit vier Pferden und zwei Personen die ganze Woche</rs></rs>
```

GrB Grundbesitzverhältnis

Markieren von Grundbesitzverhältnissen.

Protokoll F. Niemeyer: "Oft wird als Eigentümer eine Person genannt, hier wird dann geprüft, ob es sich um einen Bürger oder Adligen handelt, dementsprechend als bürgerlicher Besitz oder ritterschaftlicher Besitz markiert. Desgleichen gilt immer bei der Umschreibung von Besitzverhältnissen (z.B. gehört der Kirche = Kirchenbesitz) **Insgesamt wird beim Besitz auf die wenigen Rubriken verwiesen, die sich bereits in der Datenbank befinden.**"

Vorgehensweise in XML: Originalbegriff wird markiert und in der Datenbank normalisiert.

LN Landnutzung

Markieren von Sachbegriffen bzgl. Landnutzung.

G.M Betriebsform, Gebäude und Monumente

Auszeichnen von Betriebsformen, Gebäuden und Monumenten.

Protokoll F. Niemeyer: "Gebäude wie Mühle am Wasser werden als Wassermühle eingetragen"

Vorgehensweise in XML: Originalbegriff wird markiert und in der Datenbank normalisiert.

Tier Tiere

Markieren von Sachbegriffen bzgl. Tiervorkommen.

Protokoll F. Niemeyer: "Der Begriff Vieh, weil ähnlich unspezifisch wie Winter- und Sommergetreide wird nicht eingetragen (kommt auch überall vor !)"

Pfl. Pflanzen

Markieren von Sachbegriffen bzgl. Pflanzenvorkommen

Protokoll F. Niemeyer:

"- Vereinfachung der Pflanzenarten : z.B. Birken und Birkenbusch, Birkenbaum etc. bekommen den Eintrag „Birken“.

- Gras wird generell nicht markiert (kommt überall vor und bringt für die Auswertung nichts
- Sommer- und Wintergetreide werden nach Rücksprache mit den Editoren nicht markiert"

L.M. Solist Landschaftliche Merkmale, Sonstiges

- Erlenbruch, Kiefernbruch etc. werden als zwei Begriffe markiert : a.) Pflanze (Erle), b.) Landschaftliches Merkmal (Bruch): in xml wird dies realisiert durch die verschachtelte Markierung, die in der Datenbank entsprechend der Kategorien normalisiert wird. ([siehe oben](#))

Protokoll F. Niemeyer:

"- Begriffe wie z.B. „Sandhügel, sandiger Hügel, Hügel aus Sand etc. werden Als „Sandhügel“ eingetragen (ähnlich bei Grashügel, grasbedeckter Hügel)"

Vorgehensweise in XML: Originalbegriff wird markiert und in der Datenbank normalisiert.

c) Normalisierung der Sachbegriffe




Nachdem eine Datei bzgl. der Sachbegriffe fertig gestellt wurde können die Original-Daten in die Datenbank überspielt werden. Diese müssen dann in der Datenbank normalisiert werden.

-> näheres siehe dazu im Protokoll von Christian Kersten "[Protokoll automVortaggen.doc](#)" im gleichen Verzeichnis.

Bearbeitung der Personen


Das Ausweisen der Textabschnitte bzgl. der Personenangaben umfasst alle Vorkommen erwähnter Personen in einem Text. Bis zu vier verschiedene Angaben zu Personen können momentan erfasst werden: Vorname, Nachname, Titel und Beruf einer Person. Desweiteren wird unterschieden, ob eine Person ortsansässig oder ortsfremd ist und ob der Name innerhalb der Ortsbeschreibung bereits erwähnt wurde, es sich also um die Wiederholung einer bereits genannten Person (innerhalb einer Ortsbeschreibung) handelt. Neben dem Markieren der

Texte müssen die einzelnen Elemente ebenfalls, wie bei den GeoObjekten, zueinander in Beziehung gesetzt werden um aus den Einzelementen zusammengehörende Datensätze für die Datenbank generieren zu können. Das zentrale Element ist dabei der Nachname, welcher als Attribut eine für eine Person eindeutige ID zugewiesen bekommt.

a) , ,  *Nachname einer Person*

Wie bereits angedeutet ist der Nachname das zentrale Element für die Bestimmung eines Personendatensatzes, denn es enthält als einziges der Personen-Elemente das Attribut "xml:id". Darin wird einer Person eine eindeutige ID zugewiesen. Bei fehlendem Nachnamen kann ein leeres Nachnamens-Tag mit ID generiert werden, worauf sich dann die übrigen vorhandenen Elemente einer Person beziehen können.


Die xml:id für Personen setzt sich zusammen aus der Zeichenkette "pers" gefolgt von einer innerhalb eines Textes fortlaufenden Nummer. Bsp. für die erste Person: "pers1", für eine weitere Person: "pers2", usw.

Sich wiederholende Personennamen innerhalb eines Textes können und müssen dabei unterschiedliche xml:id-Nummer aufweisen, da jeder genannte Personenne unabhangig von seiner Hufigkeit als eigener Datensatz in die Datenbank ubernommen werden soll. Bei mehrfach Erwahnungen einer Personen, auch bei unterschiedlicher Schreibweise ist beim Taggen auf die Wahl der richtigen Buttons fur den Nachnamen zu achten! (siehe unten: )


In der Menuleiste gibt es drei verschiedene Buttons fur Nachnamen:

Zum einen bestimmen wir die Ortsansassigkeit bzw. Ortsfremdheit einer Person zum vermessen/beschriebenen Ort in der sie genannt wird. Diese Eigenschaft wird ebenfalls ausschlielich im Element fur den Nachnamen im Attribut "type" festgehalten. Fur den

Nachnamen ortsansassiger Personen wird der Button  verwendet, wohingegen Nachnamen ortsfremder Personen mit dem Button  markiert werden.

Sollte es sich bei dem zu markierenden Nachnamen einer Person um eine Wiederholung einer bereits im Text fruher erwahnten Person handeln, so wird der Button  verwendet. Wie bereits oben erwahnt, bekommt dieser Nachname eine eigene noch nicht vergebene xml:id zugeordnet und im Attribut "type" wird durch den Eintrag "repetiton" definiert, dass es sich um eine Wiederholung handelt. Zusatzlich enthalt das Element aber auch im Attribut "ref" eine Referenz auf die xml:id des Nachnamens der gleichen Person, die bereits fruher mit dem Attribut der Ortsansassigkeit bzw. Ortsfremdheit gekennzeichnet wurde. Dabei setzt sich der Eintrag fur das Attribut "ref" zusammen aus dem Symbol "#" gefolgt von der xml-id, worauf sich die Referenz bezieht. Bsp.: "#pers4". (Siehe auch die Markierung des Namens in der Beispiel-xml-Datei: "Jochim Peters")

b)  *Vorname einer Person*

Der/die Vornamen einer Person werden mit Hilfe des Buttons  markiert. Um den Vornamen einer Person zuzuordnen wird in das Attribut "ref" die "xml:id" ([siehe Nachname](#)) der Person eingetragen mit vorangestelltem "#"-Zeichen. Bsp.: #pers2

Hat eine Person mehrere hintereinander stehende Vornamen, so konnen diese zusammen in einer Klammer getaggt werden.

c) **Tit** *Titel einer Person*

Der/die Titel einer Person werden mit Hilfe des Buttons **Tit** markiert. Um den Titel einer Person zuzuordnen wird in das Attribut "ref" die "xml:id" ([siehe Nachname](#)) der Person eingetragen mit vorangestelltem "#"-Zeichen. Bsp.: #pers2

Hat eine Person mehrere hintereinander stehende Titel, so können diese zusammen in einer Klammer getaggt werden.

d) **B** *Beruf einer Person*

Der Beruf einer Person wird mit Hilfe des Buttons **B** markiert. Um den Beruf einer Person zuzuordnen wird in das Attribut "ref" die "xml:id" ([siehe Nachname](#)) der Person eingetragen mit vorangestelltem "#"-Zeichen. Bsp.: #pers2

Hat eine Person mehrere hintereinander stehende Berufe, so werden diese einzeln markiert und der Person zugeordnet. Sollte ein im Text erwähnter Beruf keiner konkreten Person zugeordnet sein, dann wird er als Beruf markiert und das Attribut "ref" gelöscht.

Bearbeitung der Orte

Alle Ortsnamen, die innerhalb eines Textes Erwähnung finden, werden für die Registersuche gekennzeichnet. Dabei werden die Orte zwei verschiedenen Gruppen zugeordnet: entweder sind es Orte, die innerhalb Schwedisch Pommerns liegen oder Orte außerhalb Schwedisch Pommerns.

In den Texten der Übersetzungen werden die Ortsnamen in der heutigen Schreibweise wiedergegeben, so dass bei evtl. Abweichungen der Ortsname korrigiert werden müsste. Handelt es sich jedoch um altdeutsche Texte bzw. Textstellen, so bleibt die Originalschreibweise des Ortsnamen im Text erhalten.


a) **Oin** *Ort in Schwedisch Pommern*

Bei der Kennzeichnung der innerhalb Schwedisch Pommerns liegenden Orte für die Registersuche können die Inhalte der Datenbank genutzt werden. Um den Arbeitsaufwand zu verringern und um Fehlern vorzubeugen, sollte die xml-Datei automatisch mit den Datenbankinhalten abgeglichen und (vor)getaggt werden: näheres siehe dazu im Protokoll "[Protokoll automVortaggen.doc](#)" im gleichen Verzeichnis.

Ein Ort, der innerhalb Schwedisch Pommerns liegt, bekommt das Attribut "key" zugeordnet in dem nach der Zeichenkette "OrtsID_" die ID des Ortes aus der Datenbank angehängt wird. Bsp.: für den Ort Konerow: key = OrtsID_697. Beim Vortaggen der Datei wird der Eintrag des Attributwertes automatisch in die xml-Datei übertragen, jedoch nur für die Orte, die ihrem Namen nach eindeutig sind. Bei Ortsnamen die nicht eindeutig einem bestimmten Ort zugeordnet werden können (z.B. Neuendorf) wird beim Vortaggen der Attributwert "OrtsID_[n]" eingetragen, wobei die Zeichenkette nach dem Unterstrich per Hand durch die richtige OrtsID aus der Datenbank ergänzt werden müsste. Fehlende Markierungen von Orten müssten nachgezeichnet werden.

Wie generell bei der Kennzeichnung von Begriffen werden auch die Ortsnamen als gesamtes Wort gekennzeichnet inklusive von Deklinations-Endungen. Wie bei den Sachbegriffen werden die Ortsnamen dann in der Datenbank normalisiert. ([siehe Sachbegriffe](#))

b)  Ort außerhalb Schwedisch Pommerns

Orte, die außerhalb Schwedisch Pommerns liegen, werden unter Verwendung des Buttons  gekennzeichnet. Diesen Ortsnamen wird im Gegensatz zu den schwedisch pommerschen Orten kein "key" zugeordnet.