



《蒙古秘史》全文检索系统

使用说明
(Python版)



金罡

呼和浩特民族学院
计算机科学与信息工程学院

2019 年 12月

《蒙古秘史》全文检索系统使用说明

一、首页

1. 网站地址：<http://www.mnuuts.com>
2. 进入网站后点击网页头部的【注册】按钮进行注册。注册后可下载检索结果数据（可不注册直接进入检索系统）。



3. 点击【进入检索系统】按钮进入系统页面。



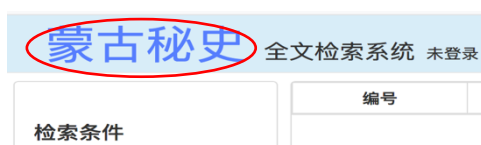
4. 首页尾部有使用说明及数据标签（tag）说明，也有一些其他可利用资源。



二、检索系统界面

1. 检索界面

- (1) 点击【进入检索系统】按钮进入检索系统界面。进入检索页面后点击左上角蓝色的【蒙古秘史】字样可返回网站首页。



- (2) 检索系统有三大功能区。一是左侧检索功能区，二是右上侧检索结果（kwic）显示区，三是右下侧文本显示区。

蒙古秘史 全文检索系统 未登录 导出检索结果

编号	前置语境	检索	后置语境	出处
检索结果显示区				

检索条件

文本

☐ 正则表达式
☐ 检索汉字

标签

词元

旁译

语义

检索

拉丁字转写文本 蒙古文本 图片

文本显示区

2. 检索功能

检索功能包括字符串检索、正则表达式检索、词性词法标签检索、词元检索、旁译检索、语义检索以及复合式检索等七种检索方法。

(1) 字符串检索

字符串检索包括拉丁字检索和汉字检索。在检索功能区中的【文本】窗体中输入拉丁字转写秘史的单词或任意字符串，按下【检索】按钮。如有需要检索汉字转写秘史单词，则在选项中勾选【检索汉字】选项后再【文本】输入框中输入汉字。

例：输入拉丁字转写单词【amin】（命）

例：输入汉字转写单词【阿敏】（命）

检索条件

文本

☐ 正则表达式
☐ 检索汉字

检索条件

文本

☐ 正则表达式
☒ 检索汉字

(2) 正则表达式检索

选择【正则表达式】选项，可利用正则表达式表示检索条件。

例如：

检索条件：以 a 开头，以 n 结尾的所有单词

正则表达式：【\sa\S+n\s】 ※ 【】符号里边的正则表达式

检索条件

文本

☒ 正则表达式

(3) 词性词法检索

在【标签】窗体中输入需要检索的词性或词法。[词性词法标签的详情请参照首页【相关说明】中的【词性词法标签】文件。](#)

例如：

在【标签】窗体中输入【PSN】，可检索到《蒙古秘史》中的所有人名。

标签	PSN
----	-----

(4) 词元检索

在【词元】窗体中输入相关词元。词元大致等于现代蒙古语单词的词根。

例如：

输入【yabu】(走)，可检索到《蒙古秘史》中以【yabu】为词元的所有单词，如 yabutuqai、yabujuǰ、yabuqui-tur 等等。

词元	yabu
----	------

(5) 旁译检索

在【旁译】窗体中输入相关旁译。

例如：

《蒙古秘史》旁译中带【馬】字的所有单词。

旁译	馬
----	---

(6) 语义检索

在【语义】窗体中输入相关语义标签。[语义标签的详情请参照首页【相关说明】中的【语义标签】文件。](#)

例如：

输入【NDIB】可检索到《蒙古秘史》中语义相关【战争】的所有单词。

语义	NDIB
----	------

(7) 复合式检索

复合式检索是指将上述 6 种条件复合为一个检索条件的复杂检索方法。

例如：检索条件：以【jü 结尾】的所有【动词】中语义相关【战争】的单词。

文本	\s\S+=jü
	☒ 正则表达式
标签	V
词元	
旁译	
语义	NDIB

3. 检索结果的利用

检索后的结果以 kwic (key word in context) 格式输出到【检索结果显示区】。kwic 格式即为被检索关键词（字符串）在中，关键词前后语境在两旁的排列格式。通过 kwic 格式可轻松观察被检索关键词（字符串）的使用语境、词形、出处以及出现频度等，并且可通过排序功能可快速查看语境特征等。

编号	前置语境	检索	后置语境	出处
1	qoričar=mergen-nü kö'ün a'ujam=boro'ul	a'ujam=boro'ul-un	kö'ün	\$002 01:01:10
2	ügüle=rün tede newü=jü	ayisu=qun	irgen-nü dotora	\$006 01:04:03
3	aldar nere yeketei	alan	=qo'a neretei gü'ün-ne ber ökte=ei üdü'üi	\$007 01:04:10
4		ariq=usun-n	a qorilartai=mergen-nü barqujin=qo'a-ča	\$008 01:05:06
5	töre=[k]sen	alan	=qo'a neretei ökin tere	\$008 01:05:07
6	burqan=bosqaqsan šinči=bayyan uriangqai-tur newü=jü	ayisu=n	a=ju'u	\$009 01:06:05
7	qori=tümed-ün qorilartai=mergen-nü ökin	ariq=usun-n	a	\$009 01:06:06

(1) 被检索单词（字符串）的频度

在 kiwc 窗体的上方记录被检索单词（字符串）在《蒙古秘史》中的使用频度。

29387个元素，543条结果				导出检索结果
前置语境	检索	后置语境	出处	
qoričar=mergen-nü kö'ün a'ujam=boro'ul	a'ujam=boro'ul-un	kö'ün	\$002 01:01:10	
ügüle=rün tede newü=jü	ayisu=qun	irgen-nü dotora	\$006 01:04:03	
aldar nere yeketei	alan	=qo'a neretei gü'ün-ne ber ökte=ei üdü'üi	\$007 01:04:10	
	ariq=usun-n	a qorilartai=mergen-nü barqujin=qo'a-ča	\$008 01:05:06	

(2) 检索结果的排序

检索结果可用四种排序方法排序。

- 词首升序，即以【a-z】的顺序排列。
- 词首降序，即以【z-a】的顺序排列。
- 词尾升序，以字符串的倒序为基准，以【a-z】的顺序排列。
- 词尾降序，以字符串的倒序为基准，以【z-a】的顺序排列。

例如：

以【检索对象】为基准，用【词尾降序】方法排序。

步骤：点击【检索对象】标签后的三角按钮，选择【词尾升序】选项。

检索				
词首升序 词首降序 词尾升序 词尾降序				
20678	hula'an^buqa-yin yabu=qsan mör-iyer	mör-tür	bayyi=qsan	\$240 10:19:08
20679	qara teri'ü man-u qajar	kösör-tür	gē=jü ot=	\$123 03:44:01
20680	tala mawui ese ke'e=be je bi	ebür-tür		\$208 08:46:03
20681	bawu=lu'a	üdü-tür	ündür-ün a=qun^ü hodun-nača olon	\$193 07:25:03
20682	üges-i an-u medere'ül=jü	üges-tür	an-u gürge=n	\$136 04:20:09

例：以【词尾降序】排序前置语境

104	ke'e=jü talbi=jü ilé=jü	ba	büşire=n güčü ök=sü ke'e=n	§149 05:06:06
105	ke'e=jü tebči=n yada=jü talbi=jü ilé=jü	širgötü		§220 09:27:09
106	yada=jü talbi=jü ilé=jü	činggis	qahan-na güčü ök=sü ke'e=n	§149 05:07:04
107	quduqul=nu'u-dača talbi=jü ilé=jü	müt	širgü'etü	§149 05:06:09
108	talbi=jü ilé=jü	ba	činggis qa'an-a güčü ög=ü=re ire=be	§220 09:28:03

以【后置语境】为基准，用【词首升序】方法排序：选择排序条件，再点击蓝色【后置语境】标签。

lū min-u altan amin saki=mui	šibawula=n	apala=n
	ke'e=n	ala=qu gü ala=asu teki mün gü alaqa=qu [^] gü bi
büy=yü	ala=n	bara=asu üktü=ksen amin ügei beye min-u
dabqur qaqa qaqa yabudaltan	ke'e=jü	bidan-ača
basa činggis qahan	ügüle=rün	bidan-u beye čerik ese
ide'en-i kebte'ül	daruqala=jü	bolqa=tuqai umdān ide'en-i
	nayita=jü	čerik qarqa=qun čerik mede=gü čerbin
qor qomsa	bol=u=asu	daruqalaqa=qsat kebte'ül-eče
geretü bü=le'e	ke'e=n	ejen ügei nuntuq-čur qočor=ču
tüge'eldü=tügei aqtas-ača	asara=jü	hǎ'ōšin ači=jü
gergei in-u qo'ai=maral a=jǎ'ai tenggis	ketül=jü	ire=be

同样的方法也可以【出处】为基准排序。

(3) 检索结果的导出

检索结果可导出为文本文件，后缀为.kwic，利用 Excel 或文本编辑器等其他工具做更深层次的统计与分析。导出结果按系统界面右方【导出检索结果】按钮。**注意：注册后方可导出检索结果。**

导出检索结果		
ba	büşire=n güčü ök=sü ke'e=n	§149 05:06:06
širgötü		§220 09:27:09

※导出的结果可能会被本地杀毒软件或浏览器屏蔽，可根据提示解除屏蔽下载使用。

4. 文本显示区的利用

系统界面右下侧即为文本显示区。文本显示区的主要功能是显示 kwic 列表中任意一个检索结果在《蒙古秘史》所在页面的文本内容。kwic 列表只显示检索对象的前后语境，而文本显示区中则显示该检索对象所在的整个页面。因此通过文本显示区可更加广面地了解该检索对象所出现的语境。

使用方法：

点击需查看的 kwic 结果行的任意一处。被点击的行变成蓝色。

编号	前置语境	检索	后置语境	出处
1	onan=müren-nū teri'ün-e	burqan=qaldun-na	nuntuqla=ju	\$001 01:01:05
2		sali=qaca'u	sali=qaca'u-yin kō'ün yeke=nidün yeke=nidün-nū kō'ün sem=soči	\$002 01:02:01
3	sali=qaca'u	sali=qaca'u-yin	kō'ün yeke=nidün yeke=nidün-nū kō'ün sem=soči	\$002 01:02:01
4		boroldai=suyalbi	jala'utu dayir boro qoyar külü'üt aqtastu	\$003 01:02:08

在文本显示区的【拉丁字转写文本】区自动显示该检索对象所在的《蒙古秘史》文本页面，显示行号、拉丁字转写文本以及汉字音写文本，并且会以蓝色显著化所在的行及单词。

拉丁字转写文本	蒙古语文本	图片
01:01:04	gergei in-u qo'ai=maral a=ji'ai tenggis ketül=jü ire=be 格兒該 亦訥 中豁埃_馬舌蘭勒 阿只埃 騰汲思 客禿勒周 亦舌列罷	
01:01:05	onan=müren-nü teri'ün-e burqan=qaldun-na nuntuqla=jü 幹難_沐舌漣訥 帖舌里兀捏 不岫罕_哈勒敦納 嫩禿黑刺周	
01:01:06	töre=ksen batačiqan a=ju'u 脫舌列克先 巴塔赤罕 阿主兀	

点击【蒙古文文本】，可显示该页面对应的蒙古文文本。

编号	前置语境	检索对象	后置语境	出处
1	qorilartai=mergen qori=tümed-ün	qajar-tur	-lyan buluqan keremün	\$009 01:06:01
2	burqan=bosqaqsan šinči=bayyan	uriangqai-tur	newü=jü ayisu=n a=ju'u	\$009 01:06:05
3	ide=küi ide'en ügei a=run činč-yin	qun-tur	qorqa=qsan görö'esün	\$026 01:16:01
4	qabur bol=ba noqot ire=küi	čaq-tur	qarčiqai-ba'an teyile'ül=jü	\$027 01:17:02

拉丁字转写文本	蒙古文文本	图片
ھولوس ەس ەمەسە ، ئۇيغۇر ۋەستەق ئۇيغۇر مسلم ەسەن ەلەم ەفەستەم ەتەستەس	ھولوس ەس ەمەسە ، ئۇيغۇر ۋەستەق ئۇيغۇر مسلم ەسەن ەلەم ەفەستەم ەتەستەس	

点击【图片】，可显示该页面对应的《蒙古秘史》影印图片。

编号	前置语境	检索对象	后置语境	出处
1	qorlartai=mergen qori=tümed-ün	qajar-tur	-lyan buluqan keremün	\$009 01:06:01
2	burqan=bosqaqsan šinči=bayyan	uriangqai-tur	newū=jū ayisu=n a=jū'u	\$009 01:06:05
3	ide=kūi ide'en ügel a=run činō-yin	qun-tur	qorqa=qsan gōrō'esün	\$026 01:16:07
4	qabur bol=ba noqot ire=kūi	čač-tur	qarčiqaī-ba'an teyile'ül=jū	\$027 01:17:02

拉丁字转写文本

蒙古文本

图片

三、汉字检索功能补充说明

利用汉字检索功能可检索《蒙古秘史》汉字音写文本。利用检索结果可进一步深入探索秘史中使用的汉字音与蒙古语音的对应关系以及汉字使用规则等。

检索例子（一）：

条件：检索所有下标小字
步骤：

- 在检索条件选项中同时选择【正则表达式】选项和【汉字检索】选项。
- 在【文本】框中输入正则表达式【_.】后按【检索】按钮。

检索条件

文本

_.

☒ 正则表达式

☒ 检索汉字

检索结果：

编号	前置语境	检索对象	后置语境	出处
1	迭額 ^{ᠳᠡᠭᠡᠨ} 騰格 ^{ᠲᠡᠭᠦᠭᠡ} 額 ^{ᠨᠡᠯᠦ} 札牙阿 ^{ᠵᠠᠶᠠᠠᠨ} 脫 ^{ᠳᠡᠯᠦ} 列	克	先 李兒帖 ^{ᠯᠢᠶᠢᠲᠦ} 赤那 ^{ᠴᠢᠨᠠ} 阿主兀	\$001 01:01:03
2	格兒該 ^{ᠭᠡᠢᠭᠡᠢ} 亦訥 ^{ᠶᠡᠨᠡᠳᠦ} 中豁埃 ^{ᠵᠣᠬᠡᠠᠢ} 馬 ^{ᠮᠠ} 闌 ^{ᠯᠠᠨ}	勒	阿只埃 ^{ᠠᠵᠢᠠᠢ} 騰汲思 ^{ᠲᠡᠭᠳᠢᠰᠢ} 客禿 ^{ᠬᠡᠲᠦ} 周 ^{ᠵᠣ} 亦 ^{ᠶᠡ} 列 ^{ᠯᠡ} 罷 ^{ᠪᠠ}	\$001 01:01:04
3	格兒該 ^{ᠭᠡᠢᠭᠡᠢ} 亦訥 ^{ᠶᠡᠨᠡᠳᠦ} 中豁埃 ^{ᠵᠣᠬᠡᠠᠢ} 馬 ^{ᠮᠠ} 闌 ^{ᠯᠠᠨ} 阿只埃 ^{ᠠᠵᠢᠠᠢ} 騰汲思 ^{ᠲᠡᠭᠳᠢᠰᠢ} 客禿 ^{ᠬᠡᠲᠦ}	勒	周 ^{ᠵᠣ} 亦 ^{ᠶᠡ} 列 ^{ᠯᠡ} 罷 ^{ᠪᠠ}	\$001 01:01:04
4	幹難 ^{ᠭᠠᠨᠠᠨ} 沐 ^{ᠮᠤ} 達訥 ^{ᠳᠠᠨᠡ} 帖 ^{ᠲᠦ} 里兀捏 ^{ᠯᠢᠦᠨᠡ} 不 ^{ᠨᠠ} 峇罕 ^{ᠪᠠᠬᠠᠨ} 哈	勒	敦納 ^{ᠳᠤᠨᠠ} 嫩禿 ^{ᠨᠡᠳᠦ} 刺 ^{ᠰᠢ} 周 ^{ᠵᠣ}	\$001 01:01:05
5	幹難 ^{ᠭᠠᠨᠠᠨ} 沐 ^{ᠮᠤ} 達訥 ^{ᠳᠠᠨᠡ} 帖 ^{ᠲᠦ} 里兀捏 ^{ᠯᠢᠦᠨᠡ} 不 ^{ᠨᠠ} 峇罕 ^{ᠪᠠᠬᠠᠨ} 哈 ^{ᠬᠠ} 敦納 ^{ᠳᠤᠨᠠ} 嫩禿 ^{ᠨᠡᠳᠦ}	黑	刺 ^{ᠰᠢ} 周 ^{ᠵᠣ}	\$001 01:01:05