

# Web genre classification with methods for structured output prediction<sup>1</sup>

Gjorgji Madjarov<sup>a</sup>, Vedrana Vidulin<sup>b</sup>, Ivica Dimitrovski<sup>a</sup>, Dragi Kocev<sup>c</sup>

<sup>a</sup>*Faculty of Computer Science and Engineering, University Ss. Cyril and Methodius, Skopje, Macedonia*

<sup>b</sup>*Ruđer Bošković Institute, Zagreb, Croatia*

<sup>c</sup>*Department of Knowledge Technologies, Jožef Stefan Institute, Ljubljana, Slovenia*

---

## Abstract

This is the supplementary information for the paper. It contains all of the results obtained with the experimental evaluation using bagging.

**Keywords:** web genre classification; hierarchy construction; hierarchical multi-label classification

---

## 1. Complete results using bagging

### Acknowledgments

We acknowledge the financial support of the European Commission through the grant ICT-2013-612944 MAESTRA.

---

<sup>1</sup>The first two authors should be regarded as joint first authors.

Email addresses: Gjorgji.Madjarov@finki.ukim.mk (Gjorgji Madjarov), Vedrana.Vidulin@irb.hr (Vedrana Vidulin), Ivica.Dimitrovski@finki.ukim.mk (Ivica Dimitrovski), Dragi.Kocev@ijs.si (Dragi Kocev)

Table 1: The performance of the different machine learning tasks applied on the different features using bagging. The hierarchy construction method are abbreviated as follows: balanced  $k$ -means clustering (BkM), predictive clustering tree (PCT), clustering with complete linkage (CL), clustering with single linkage (SL), random (RND) and manual (MAN). The evaluation measure is *Hamming loss* where smaller value means better result.

20-genres	HMC												MLC		SC		
	CL	SL	B2M	B3M	B4M	PCT	RND	MAN	CL	SL	B2M	B3M	B4M	PCT	RND	MAN	
BOWSrf	0.07	0.07	0.068	0.069	0.067	0.068	0.071	0.067	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.062	0.067
TFIDFSrf	0.069	0.07	0.067	0.067	0.066	0.067	0.069	0.065	0.002	0.001	0.002	0.002	0.001	0.001	0.001	0.061	0.067
Struc	0.083	0.082	0.081	0.081	0.08	0.082	0.08	0.082	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.075	0.087
Pres	0.084	0.085	0.086	0.085	0.085	0.085	0.086	0.086	0.001	0.001	0.001	0.001	0.001	0.002	0.001	0.082	0.086
Context	0.097	0.098	0.096	0.095	0.097	0.095	0.097	0.096	0.02	0.019	0.019	0.019	0.019	0.019	0.019	0.092	0.097
PV_R	0.076	0.099	0.071	0.071	0.072	0.071	0.075	0.07	0.001	0.001	0	0	0	0.001	0	0.083	0.085
PV_C	0.076	0.079	0.069	0.069	0.069	0.069	0.075	0.07	0	0	0	0	0	0	0	0.081	0.082
PV_GLR	0.102	0.104	0.097	0.096	0.095	0.097	0.1	0.095	0	0.001	0.001	0	0	0	0	0.084	0.092
PV_GLC	0.08	0.08	0.073	0.075	0.074	0.075	0.081	0.075	0	0	0	0	0	0	0	0.089	0.091
Ngrams	0.075	0.073	0.071	0.072	0.072	0.071	0.074	0.07	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.067	0.071
SPIRIT																	
BOWSrf	0.087	0.088	0.084	0.087	0.083	0.083	0.088	0.083	0.008	0.007	0.008	0.007	0.007	0.008	0.008	0.087	0.086
TFIDFSrf	0.087	0.085	0.081	0.082	0.081	0.083	0.087	0.082	0.006	0.006	0.006	0.006	0.006	0.005	0.006	0.087	0.086
Struc	0.086	0.089	0.085	0.086	0.087	0.088	0.087	0.087	0.009	0.008	0.009	0.008	0.008	0.008	0.008	0.086	0.087
Pres	0.099	0.101	0.101	0.102	0.1	0.101	0.101	0.101	0.009	0.009	0.011	0.01	0.009	0.009	0.011	0.102	0.101
Context	0.117	0.117	0.119	0.118	0.119	0.118	0.118	0.118	0.039	0.037	0.04	0.038	0.037	0.037	0.039	0.122	0.121
PV_R	0.099	0.098	0.089	0.088	0.091	0.091	0.092	0.09	0.005	0.005	0.005	0.004	0.004	0.005	0.005	0.094	0.095
PV_C	0.094	0.094	0.093	0.094	0.093	0.091	0.095	0.094	0.005	0.004	0.006	0.005	0.005	0.007	0.005	0.094	0.092
PV_GLR	0.103	0.104	0.105	0.105	0.103	0.106	0.105	0.004	0.003	0.004	0.003	0.004	0.003	0.004	0.004	0.102	0.099
PV_GLC	0.098	0.098	0.097	0.096	0.099	0.098	0.1	0.097	0.005	0.004	0.005	0.004	0.004	0.006	0.005	0.095	0.094
Ngrams	0.094	0.094	0.094	0.093	0.093	0.095	0.093	0.094	0.005	0.004	0.005	0.005	0.005	0.004	0.005	0.094	0.096

Table 2: The performance of the different machine learning tasks applied on the different features using bagging. The hierarchy construction method are abbreviated as follows: balanced  $k$ -means clustering (BkM), predictive clustering tree (PCT), clustering with complete linkage (CL), clustering with single linkage (SL), random (RND) and manual (MAN). The evaluation measure is accuracy where larger value means better result.

20-genres	HMC						HSC						MLC	SC		
	CL	SL	B2M	B3M	B4M	PCT	RND	MAN	CL	SL	B2M	B3M	B4M	PCT	RND	MAN
BOWSrf	0.386	0.38	0.454	0.443	0.45	0.443	0.405	0.451	0.988	0.989	0.987	0.99	0.985	0.988	0.468	0.374
TFIDFSrf	0.407	0.396	0.459	0.459	0.462	0.459	0.417	0.467	0.975	0.986	0.985	0.98	0.988	0.988	0.487	0.393
Struc	0.271	0.273	0.306	0.311	0.313	0.318	0.289	0.321	0.985	0.991	0.991	0.988	0.986	0.986	0.296	0.204
Pres	0.285	0.279	0.308	0.317	0.305	0.31	0.298	0.307	0.986	0.984	0.984	0.989	0.986	0.982	0.287	0.211
Context	0.21	0.206	0.242	0.244	0.235	0.241	0.222	0.24	0.817	0.821	0.819	0.821	0.821	0.82	0.819	0.231
PV_R	0.243	0.274	0.313	0.312	0.302	0.3	0.253	0.314	0.994	0.994	0.995	0.995	0.994	0.994	0.996	0.32
PV_C	0.295	0.279	0.384	0.382	0.378	0.382	0.319	0.38	0.997	0.997	0.996	0.995	0.995	0.997	0.996	0.29
PV_GI_R	0.246	0.237	0.304	0.292	0.301	0.302	0.262	0.297	0.996	0.994	0.994	0.997	0.993	0.996	0.995	0.282
PV_GI_C	0.257	0.255	0.309	0.305	0.302	0.302	0.253	0.303	0.995	0.996	0.996	0.998	0.996	0.996	0.995	0.181
Ngrams	0.334	0.341	0.421	0.413	0.414	0.417	0.359	0.425	0.993	0.993	0.994	0.993	0.993	0.993	0.994	0.36
<hr/>																
SPiRiT																
BOWSrf	0.541	0.542	0.538	0.543	0.539	0.547	0.537	0.546	0.95	0.953	0.949	0.954	0.955	0.957	0.946	0.543
TFIDFSrf	0.546	0.554	0.546	0.551	0.557	0.547	0.549	0.552	0.959	0.96	0.959	0.957	0.963	0.963	0.952	0.528
Struc	0.532	0.526	0.539	0.535	0.535	0.532	0.533	0.537	0.934	0.938	0.941	0.936	0.941	0.946	0.946	0.502
Pres	0.483	0.478	0.478	0.473	0.482	0.479	0.481	0.482	0.933	0.934	0.929	0.933	0.935	0.933	0.926	0.444
Context	0.392	0.395	0.394	0.394	0.392	0.393	0.391	0.394	0.771	0.781	0.767	0.775	0.78	0.766	0.778	0.362
PV_R	0.508	0.514	0.521	0.521	0.511	0.512	0.502	0.519	0.962	0.962	0.963	0.964	0.967	0.967	0.957	0.495
PV_C	0.493	0.496	0.501	0.507	0.508	0.515	0.5	0.501	0.964	0.97	0.957	0.965	0.965	0.948	0.959	0.491
PV_GI_R	0.482	0.478	0.481	0.476	0.477	0.486	0.477	0.48	0.967	0.971	0.967	0.974	0.97	0.972	0.969	0.458
PV_GI_C	0.482	0.481	0.488	0.497	0.482	0.492	0.481	0.492	0.962	0.973	0.965	0.969	0.968	0.969	0.959	0.468
Ngrams	0.503	0.501	0.506	0.509	0.513	0.51	0.503	0.511	0.966	0.972	0.967	0.967	0.968	0.967	0.966	0.507

Table 3: The performance of the different machine learning tasks applied on the different features using bagging. The hierarchy construction method are abbreviated as follows: balanced  $k$ -means clustering (BkM), predictive clustering tree (PCT), clustering with complete linkage (CL), clustering with single linkage (SL), random (RND) and manual (MAN). The evaluation measure is *precision* where larger value means better result.

20-genres	HMC										HSC				MLC		SC	
	CL	SL	B2M	B3M	B4M	PCT	RND	MAN	CL	SL	B2M	B3M	B4M	PCT	RND	MAN		
BOWSrf	0.436	0.432	0.503	0.495	0.501	0.493	0.462	0.504	0.991	0.991	0.991	0.99	0.993	0.989	0.991	0.52	0.422	
TFIDFSrf	0.462	0.45	0.51	0.512	0.511	0.51	0.474	0.52	0.984	0.988	0.987	0.989	0.989	0.989	0.988	0.54	0.449	
Struc	0.313	0.317	0.349	0.359	0.361	0.362	0.338	0.37	0.989	0.995	0.994	0.992	0.993	0.991	0.99	0.993	0.341	0.235
Pres	0.331	0.321	0.349	0.357	0.347	0.351	0.341	0.344	0.989	0.988	0.992	0.988	0.989	0.986	0.985	0.326	0.245	
Context	0.238	0.232	0.271	0.274	0.263	0.271	0.251	0.269	0.826	0.829	0.827	0.829	0.828	0.827	0.827	0.259	0.23	
PV_R	0.283	0.308	0.36	0.358	0.344	0.346	0.298	0.359	0.997	0.997	0.997	0.997	0.996	0.996	0.998	0.359	0.252	
PV_C	0.341	0.323	0.433	0.432	0.426	0.428	0.372	0.425	0.998	0.998	0.997	0.997	0.996	0.998	0.999	0.422	0.33	
PV_GlR	0.28	0.272	0.339	0.328	0.337	0.337	0.301	0.337	0.996	0.996	0.997	0.997	0.996	0.998	0.998	0.323	0.208	
PV_GlC	0.302	0.298	0.357	0.351	0.351	0.348	0.294	0.35	0.997	0.997	0.997	0.997	0.997	0.997	0.998	0.345	0.273	
Ngrams	0.382	0.385	0.472	0.458	0.462	0.463	0.405	0.47	0.995	0.994	0.995	0.994	0.993	0.994	0.996	0.464	0.413	
SPIRIT																		
BOWSrf	0.642	0.642	0.664	0.643	0.663	0.669	0.638	0.67	0.959	0.963	0.96	0.965	0.964	0.965	0.956	0.958	0.642	0.618
TFIDFSrf	0.649	0.658	0.666	0.67	0.666	0.657	0.644	0.667	0.972	0.975	0.974	0.974	0.974	0.976	0.972	0.968	0.649	0.618
Struc	0.646	0.638	0.656	0.65	0.648	0.643	0.642	0.646	0.953	0.955	0.959	0.957	0.96	0.96	0.956	0.957	0.652	0.61
Pres	0.586	0.579	0.577	0.576	0.584	0.581	0.583	0.584	0.947	0.948	0.946	0.952	0.949	0.948	0.944	0.946	0.586	0.55
Context	0.491	0.492	0.491	0.495	0.494	0.495	0.492	0.493	0.821	0.824	0.817	0.825	0.826	0.826	0.817	0.823	0.486	0.464
PV_R	0.596	0.604	0.635	0.638	0.626	0.628	0.617	0.633	0.97	0.971	0.972	0.975	0.977	0.978	0.968	0.969	0.622	0.588
PV_C	0.609	0.61	0.609	0.617	0.615	0.62	0.607	0.609	0.975	0.979	0.969	0.977	0.975	0.975	0.962	0.972	0.614	0.577
PV_GlR	0.591	0.592	0.585	0.583	0.587	0.593	0.583	0.582	0.975	0.978	0.978	0.985	0.981	0.981	0.978	0.978	0.587	0.571
PV_GlC	0.596	0.605	0.598	0.61	0.596	0.601	0.593	0.607	0.972	0.982	0.978	0.98	0.98	0.98	0.97	0.973	0.617	0.58
Ngrams	0.613	0.611	0.612	0.613	0.616	0.618	0.615	0.973	0.977	0.974	0.974	0.976	0.976	0.974	0.974	0.607	0.602	

Table 4: The performance of the different machine learning tasks applied on the different features using bagging. The hierarchy construction method are abbreviated as follows: balanced  $k$ -means clustering (BkM), predictive clustering tree (PCT), clustering with complete linkage (CL), clustering with single linkage (SL), random (RND) and manual (MAN). The evaluation measure is *recall* where larger value means better result.

20-genres	HMC						HSC						MLC	SC				
	CL	SL	B2M	B3M	B4M	PCT	RND	MAN	CL	SL	B2M	B3M	B4M	PCT	RND	MAN		
BOWSrf	0.474	0.471	0.569	0.56	0.57	0.566	0.504	0.569	0.994	0.995	0.995	0.992	0.995	0.995	0.995	0.562	0.488	
TFIDFSrf	0.503	0.493	0.583	0.588	0.587	0.592	0.521	0.592	0.98	0.997	0.996	0.983	0.997	0.997	0.998	0.597	0.516	
Struc	0.356	0.354	0.41	0.416	0.409	0.418	0.38	0.413	0.987	0.992	0.992	0.989	0.99	0.987	0.992	0.359	0.289	
Pres	0.361	0.361	0.401	0.415	0.399	0.404	0.376	0.401	0.992	0.989	0.989	0.993	0.991	0.987	0.987	0.351	0.291	
Context	0.264	0.259	0.309	0.309	0.302	0.31	0.281	0.301	0.89	0.891	0.89	0.893	0.892	0.891	0.891	0.282	0.293	
PV_R	0.287	0.442	0.382	0.373	0.376	0.365	0.306	0.382	0.995	0.996	0.996	0.996	0.995	0.995	0.997	0.443	0.326	
PV_C	0.36	0.353	0.473	0.473	0.46	0.474	0.402	0.473	0.997	0.998	0.997	0.996	0.997	0.998	0.999	0.554	0.442	
PV_GI_R	0.391	0.384	0.486	0.472	0.473	0.481	0.41	0.475	1	0.995	0.995	0.999	0.994	0.996	1	0.996	0.373	0.269
PV_GI_C	0.328	0.328	0.386	0.38	0.378	0.383	0.32	0.382	0.995	0.997	0.996	0.998	0.997	0.997	0.997	0.449	0.363	
Ngrams	0.416	0.419	0.531	0.52	0.52	0.527	0.443	0.533	0.998	0.999	0.998	0.998	0.998	0.998	0.999	0.505	0.488	
<b>SPIRIT</b>																		
BOWSrf	0.663	0.67	0.619	0.679	0.63	0.633	0.664	0.627	0.979	0.979	0.977	0.98	0.982	0.982	0.976	0.677	0.652	
TFIDFSrf	0.663	0.675	0.63	0.637	0.646	0.638	0.68	0.638	0.973	0.975	0.971	0.968	0.973	0.974	0.968	0.664	0.644	
Struc	0.601	0.609	0.613	0.621	0.618	0.614	0.625	0.581	0.951	0.955	0.956	0.952	0.955	0.957	0.975	0.631	0.592	
Pres	0.58	0.578	0.582	0.581	0.58	0.581	0.59	0.587	0.961	0.962	0.958	0.959	0.962	0.96	0.959	0.956	0.539	
Context	0.497	0.499	0.507	0.503	0.499	0.499	0.501	0.506	0.851	0.865	0.848	0.859	0.857	0.865	0.846	0.86	0.496	0.47
PV_R	0.642	0.643	0.592	0.601	0.59	0.591	0.583	0.593	0.978	0.977	0.974	0.977	0.978	0.977	0.974	0.971	0.659	0.616
PV_C	0.586	0.587	0.587	0.603	0.606	0.61	0.593	0.596	0.977	0.982	0.973	0.978	0.975	0.978	0.969	0.972	0.571	0.58
PV_GI_R	0.599	0.6	0.612	0.595	0.605	0.619	0.611	0.612	0.975	0.979	0.975	0.98	0.976	0.976	0.97	0.574	0.555	
PV_GI_C	0.573	0.57	0.589	0.595	0.574	0.585	0.59	0.584	0.975	0.984	0.977	0.98	0.975	0.975	0.969	0.564	0.553	
Ngrams	0.585	0.586	0.603	0.596	0.615	0.607	0.597	0.613	0.983	0.988	0.986	0.985	0.983	0.987	0.983	0.589	0.626	

Table 5: The performance of the different machine learning tasks applied on the different features using bagging. The hierarchy construction method are abbreviated as follows: balanced  $k$ -means clustering (BkM), predictive clustering tree (PCT), clustering with complete linkage (CL), clustering with single linkage (SL) and manual (MAN). The evaluation measure is  $F_1$  score where larger value means better result.

20-genres	HMC										HSC				MLC		SC	
	CL	SL	B2M	B3M	B4M	PCT	RND	MAN	CL	SL	B2M	B3M	B4M	PCT	RND	MAN		
BOWSrf	0.454	0.451	0.534	0.525	0.533	0.527	0.482	0.535	0.993	0.993	0.992	0.993	0.994	0.991	0.993	0.993	0.54	0.452
TFIDFSrf	0.482	0.471	0.544	0.547	0.547	0.548	0.496	0.554	0.982	0.993	0.991	0.986	0.993	0.993	0.993	0.993	0.567	0.48
Struc	0.333	0.334	0.377	0.386	0.384	0.388	0.358	0.39	0.988	0.993	0.993	0.991	0.991	0.991	0.989	0.992	0.35	0.26
Pres	0.345	0.34	0.373	0.384	0.371	0.376	0.358	0.371	0.99	0.989	0.989	0.993	0.988	0.99	0.986	0.986	0.338	0.266
Context	0.251	0.245	0.289	0.291	0.281	0.289	0.265	0.285	0.857	0.859	0.857	0.86	0.859	0.858	0.858	0.858	0.27	0.258
PV_R	0.285	0.364	0.371	0.365	0.359	0.355	0.302	0.371	0.996	0.996	0.997	0.996	0.996	0.996	0.996	0.998	0.397	0.285
PV_C	0.35	0.337	0.452	0.452	0.442	0.45	0.387	0.448	0.998	0.998	0.997	0.997	0.996	0.998	0.998	0.999	0.479	0.378
PV_GLR	0.326	0.319	0.4	0.387	0.394	0.397	0.347	0.395	0.998	0.995	0.996	0.998	0.994	0.996	0.998	0.997	0.346	0.235
PV_GLC	0.315	0.312	0.371	0.365	0.364	0.365	0.306	0.365	0.996	0.997	0.997	0.998	0.997	0.997	0.998	0.997	0.39	0.312
Ngrams	0.398	0.401	0.5	0.487	0.49	0.493	0.423	0.499	0.996	0.996	0.996	0.996	0.995	0.996	0.997	0.484	0.448	
SPIRIT																		
BOWSrf	0.652	0.656	0.64	0.66	0.646	0.651	0.648	0.969	0.971	0.969	0.972	0.973	0.973	0.966	0.966	0.659	0.635	
TFIDFSrf	0.656	0.666	0.648	0.653	0.656	0.647	0.662	0.652	0.973	0.975	0.972	0.971	0.973	0.975	0.97	0.966	0.664	0.63
Struc	0.623	0.623	0.634	0.635	0.633	0.63	0.628	0.635	0.952	0.955	0.957	0.955	0.957	0.959	0.965	0.966	0.641	0.601
Pres	0.583	0.579	0.579	0.578	0.582	0.581	0.587	0.586	0.954	0.955	0.952	0.955	0.955	0.956	0.954	0.951	0.587	0.544
Context	0.494	0.495	0.499	0.499	0.497	0.497	0.496	0.499	0.835	0.844	0.832	0.841	0.841	0.845	0.832	0.841	0.491	0.467
PV_R	0.618	0.623	0.613	0.619	0.607	0.609	0.6	0.613	0.974	0.974	0.973	0.976	0.978	0.978	0.971	0.97	0.64	0.602
PV_C	0.597	0.598	0.598	0.61	0.61	0.615	0.6	0.602	0.976	0.98	0.971	0.977	0.975	0.976	0.966	0.972	0.592	0.579
PV_GLR	0.595	0.596	0.589	0.596	0.606	0.596	0.596	0.975	0.979	0.976	0.982	0.979	0.978	0.977	0.974	0.58	0.563	
PV_GLC	0.584	0.587	0.593	0.602	0.585	0.593	0.588	0.599	0.973	0.983	0.978	0.98	0.979	0.98	0.972	0.971	0.589	0.566
Ngrams	0.599	0.598	0.608	0.604	0.615	0.612	0.606	0.614	0.978	0.982	0.98	0.98	0.98	0.98	0.978	0.978	0.598	0.613

Table 6: The performance of the different machine learning tasks applied on the different features using bagging. The hierarchy construction method are abbreviated as follows: balanced  $k$ -means clustering (BkM), predictive clustering tree (PCT), clustering with complete linkage (CL), clustering with single linkage (SL), random (RND) and manual (MAN). The evaluation measure is *subset accuracy* where larger value means better result.

20-genres	HMC						HSC						MLC	SC				
	CL	SL	B2M	B3M	B4M	PCT	RND	MAN	CL	SL	B2M	B3M	B4M	PCT	RND	MAN		
BOWSrf	0.251	0.24	0.292	0.277	0.283	0.273	0.254	0.283	0.977	0.975	0.974	0.981	0.972	0.977	0.326	0.229		
TFIDFSrf	0.262	0.249	0.285	0.281	0.288	0.276	0.261	0.29	0.959	0.971	0.968	0.966	0.974	0.975	0.973	0.319	0.233	
Struc	0.151	0.157	0.166	0.167	0.177	0.182	0.156	0.186	0.98	0.985	0.985	0.982	0.981	0.984	0.979	0.983	0.192	0.118
Pres	0.168	0.16	0.185	0.186	0.175	0.183	0.181	0.184	0.976	0.973	0.973	0.98	0.975	0.977	0.967	0.973	0.19	0.121
Context	0.135	0.132	0.156	0.157	0.147	0.152	0.14	0.157	0.729	0.736	0.734	0.733	0.733	0.735	0.73	0.158	0.105	
PV_R	0.16	0.096	0.199	0.208	0.186	0.191	0.16	0.202	0.989	0.99	0.991	0.99	0.991	0.992	0.988	0.993	0.166	0.113
PV_C	0.189	0.169	0.25	0.247	0.255	0.249	0.189	0.248	0.994	0.993	0.993	0.992	0.993	0.995	0.993	0.996	0.202	0.145
PV_GI_R	0.092	0.079	0.117	0.1	0.117	0.117	0.098	0.107	0.992	0.991	0.99	0.993	0.99	0.994	0.992	0.992	0.16	0.094
PV_GI_C	0.147	0.145	0.192	0.189	0.183	0.182	0.15	0.182	0.992	0.993	0.993	0.997	0.994	0.993	0.993	0.993	0.135	0.123
Ngrams	0.206	0.221	0.264	0.263	0.263	0.264	0.23	0.273	0.987	0.986	0.986	0.986	0.987	0.986	0.987	0.988	0.281	0.208
<b>SPIRIT</b>																		
BOWSrf	0.316	0.316	0.339	0.309	0.328	0.342	0.311	0.345	0.905	0.91	0.9	0.909	0.91	0.915	0.895	0.902	0.308	0.33
TFIDFSrf	0.325	0.326	0.336	0.35	0.355	0.343	0.321	0.348	0.924	0.925	0.923	0.92	0.924	0.929	0.923	0.915	0.305	0.323
Struc	0.349	0.329	0.348	0.33	0.34	0.335	0.345	0.343	0.889	0.896	0.898	0.89	0.899	0.899	0.904	0.901	0.33	0.299
Pres	0.298	0.289	0.284	0.274	0.291	0.289	0.285	0.287	0.882	0.884	0.87	0.878	0.884	0.879	0.863	0.875	0.284	0.265
Context	0.206	0.209	0.21	0.209	0.21	0.21	0.201	0.207	0.618	0.63	0.618	0.618	0.634	0.63	0.614	0.628	0.18	0.179
PV_R	0.294	0.298	0.339	0.333	0.322	0.326	0.315	0.335	0.934	0.932	0.935	0.935	0.939	0.942	0.928	0.926	0.309	0.294
PV_C	0.291	0.301	0.312	0.311	0.316	0.321	0.308	0.304	0.935	0.943	0.922	0.933	0.93	0.935	0.908	0.928	0.298	0.274
PV_GI_R	0.27	0.262	0.27	0.268	0.26	0.262	0.258	0.268	0.948	0.954	0.943	0.955	0.952	0.955	0.949	0.947	0.267	0.265
PV_GI_C	0.289	0.282	0.289	0.299	0.288	0.304	0.278	0.292	0.934	0.95	0.932	0.942	0.94	0.943	0.919	0.931	0.308	0.288
Ngrams	0.315	0.306	0.306	0.316	0.309	0.311	0.301	0.304	0.937	0.949	0.935	0.934	0.939	0.936	0.935	0.939	0.309	0.308

Table 7: The performance of the different machine learning tasks applied on the different features using bagging. The hierarchy construction method are abbreviated as follows: balanced  $k$ -means clustering (BkM), predictive clustering tree (PCT), clustering with complete linkage (CL), clustering with single linkage (SL) and manual (MAN). The evaluation measure is *micro precision* where larger value means better result.

20-genres	HMC										HSC				MLC		SC		
	CL	SL	B2M	B3M	B4M	PCT	RND	MAN	CL	SL	B2M	B3M	B4M	PCT	RND	MAN			
BOWSrf	0.475	0.475	0.493	0.486	0.499	0.492	0.472	0.499	0.989	0.989	0.987	0.989	0.992	0.988	0.989	0.537	0.499		
TFIDFSrf	0.486	0.477	0.5	0.499	0.507	0.497	0.488	0.512	0.992	0.982	0.994	0.984	0.984	0.984	0.982	0.541	0.504		
Struc	0.366	0.371	0.393	0.395	0.397	0.399	0.379	0.4	0.997	0.998	0.998	0.998	0.997	0.998	0.997	0.996	0.427	0.322	
Pres	0.368	0.357	0.364	0.37	0.368	0.368	0.363	0.365	0.99	0.991	0.991	0.992	0.992	0.989	0.991	0.377	0.332		
Context	0.265	0.26	0.287	0.291	0.282	0.293	0.272	0.283	0.826	0.833	0.831	0.829	0.833	0.831	0.829	0.292	0.277		
PV_R	0.403	0.318	0.463	0.462	0.457	0.456	0.416	0.468	0.998	0.998	0.998	0.998	0.999	0.998	0.998	0.388	0.345		
PV_C	0.419	0.396	0.481	0.487	0.485	0.481	0.435	0.477	0.999	0.998	0.998	0.998	0.997	0.999	0.998	0.417	0.396		
PV_GlR	0.293	0.287	0.338	0.337	0.34	0.334	0.305	0.342	0.994	0.999	0.995	0.998	1	0.994	0.999	0.37	0.286		
PV_GlC	0.382	0.38	0.443	0.429	0.436	0.428	0.374	0.432	0.999	0.998	0.999	0.999	0.999	0.998	0.999	0.362	0.327		
Ngrams	0.436	0.446	0.474	0.467	0.469	0.472	0.447	0.482	0.992	0.991	0.993	0.992	0.992	0.993	0.992	0.5	0.468		
SPIRIT																			
BOWSrf	0.608	0.603	0.639	0.604	0.643	0.642	0.601	0.641	0.956	0.961	0.957	0.961	0.959	0.962	0.952	0.959	0.605	0.614	
TFIDFSrf	0.609	0.618	0.648	0.646	0.646	0.638	0.604	0.644	0.979	0.979	0.979	0.982	0.98	0.982	0.983	0.979	0.605	0.618	
Struc	0.627	0.612	0.632	0.623	0.618	0.614	0.621	0.619	0.966	0.968	0.97	0.97	0.972	0.973	0.956	0.621	0.624		
Pres	0.559	0.552	0.55	0.546	0.557	0.55	0.55	0.552	0.957	0.957	0.951	0.96	0.957	0.956	0.951	0.543	0.553		
Context	0.477	0.479	0.471	0.475	0.475	0.475	0.475	0.475	0.823	0.824	0.818	0.822	0.832	0.824	0.825	0.824	0.456	0.46	
PV_R	0.553	0.559	0.615	0.617	0.603	0.601	0.599	0.61	0.977	0.979	0.983	0.983	0.984	0.986	0.977	0.98	0.573	0.573	
PV_C	0.588	0.587	0.592	0.585	0.586	0.595	0.581	0.584	0.982	0.983	0.976	0.982	0.981	0.983	0.971	0.982	0.59	0.594	
PV_GlR	0.542	0.541	0.534	0.532	0.531	0.54	0.527	0.529	0.989	0.989	0.988	0.992	0.991	0.993	0.989	0.994	0.546	0.562	
PV_GlC	0.566	0.569	0.568	0.576	0.562	0.566	0.553	0.569	0.981	0.985	0.98	0.985	0.984	0.973	0.985	0.584	0.593		
Ngrams	0.585	0.587	0.583	0.589	0.584	0.579	0.587	0.582	0.975	0.979	0.972	0.975	0.976	0.979	0.974	0.976	0.583	0.568	

Table 8: The performance of the different machine learning tasks applied on the different features using bagging. The hierarchy construction method are abbreviated as follows: balanced  $k$ -means clustering (BkM), predictive clustering tree (PCT), clustering with complete linkage (CL), clustering with single linkage (SL), random (RND) and manual (MAN). The evaluation measure is *micro recall* where larger value means better result.

20-genres	HMC						HSC						MLC		SC	
	CL	SL	B2M	B3M	B4M	PCT	RND	MAN	CL	SL	B2M	B3M	B4M	PCT	RND	MAN
BOWSrf	0.443	0.44	0.537	0.529	0.54	0.535	0.472	0.537	0.993	0.994	0.992	0.993	0.991	0.993	0.536	0.469
TFIDFSrf	0.473	0.468	0.56	0.565	0.562	0.568	0.49	0.566	0.976	0.997	0.994	0.979	0.996	0.997	0.574	0.496
Struc	0.328	0.324	0.375	0.383	0.376	0.384	0.35	0.384	0.985	0.989	0.991	0.988	0.985	0.99	0.332	0.267
Pres	0.343	0.342	0.378	0.389	0.378	0.38	0.353	0.377	0.992	0.988	0.98	0.992	0.988	0.991	0.988	0.334
Context	0.25	0.247	0.29	0.293	0.286	0.292	0.264	0.283	0.894	0.893	0.892	0.896	0.895	0.894	0.894	0.266
PV_R	0.266	0.423	0.36	0.351	0.34	0.282	0.361	0.993	0.994	0.994	0.995	0.995	0.995	0.993	0.997	0.424
PV_C	0.34	0.332	0.446	0.448	0.433	0.447	0.374	0.443	0.997	0.997	0.997	0.996	0.997	0.997	0.999	0.534
PV_GI_R	0.374	0.367	0.457	0.446	0.449	0.449	0.386	0.448	1	0.994	0.994	1	0.994	0.996	1	0.995
PV_GI_C	0.313	0.311	0.365	0.359	0.353	0.361	0.3	0.36	0.994	0.996	0.995	0.998	0.997	0.996	0.996	0.431
Ngrams	0.393	0.392	0.501	0.497	0.493	0.504	0.414	0.504	0.997	0.998	0.997	0.998	0.998	0.998	0.998	0.435
<hr/>																
SPiRiT																
BOWSrf	0.636	0.641	0.588	0.649	0.596	0.6	0.639	0.597	0.978	0.977	0.974	0.976	0.98	0.974	0.973	0.645
TFIDFSrf	0.637	0.647	0.602	0.602	0.613	0.61	0.653	0.609	0.971	0.971	0.967	0.963	0.97	0.97	0.961	0.647
Struc	0.573	0.577	0.584	0.593	0.585	0.59	0.583	0.596	0.952	0.956	0.954	0.951	0.954	0.957	0.976	0.574
Pres	0.55	0.546	0.555	0.552	0.551	0.551	0.556	0.556	0.962	0.962	0.955	0.955	0.955	0.961	0.956	0.507
Context	0.462	0.465	0.469	0.467	0.462	0.465	0.462	0.472	0.835	0.853	0.833	0.841	0.844	0.85	0.847	0.441
PV_R	0.617	0.619	0.564	0.568	0.562	0.561	0.554	0.564	0.979	0.977	0.974	0.976	0.977	0.975	0.972	0.637
PV_C	0.558	0.552	0.563	0.572	0.579	0.584	0.567	0.567	0.975	0.98	0.971	0.975	0.975	0.965	0.969	0.597
PV_GI_R	0.562	0.555	0.577	0.555	0.564	0.578	0.574	0.577	0.977	0.982	0.974	0.974	0.979	0.977	0.97	0.559
PV_GI_C	0.54	0.537	0.555	0.565	0.548	0.558	0.556	0.557	0.974	0.982	0.973	0.977	0.97	0.968	0.968	0.536
Ngrams	0.561	0.564	0.579	0.578	0.592	0.582	0.575	0.591	0.982	0.988	0.985	0.983	0.982	0.981	0.985	0.605

Table 9: The performance of the different machine learning tasks applied on the different features using bagging. The hierarchy construction method are abbreviated as follows: balanced  $k$ -means clustering (BkM), predictive clustering tree (PCT), clustering with complete linkage (CL), clustering with single linkage (SL), random (RND) and manual (MAN). The evaluation measure is *micro F<sub>1</sub>* where larger value means better result.

20-genres	HMC										HSC				MLC		SC	
	CL	SL	B2M	B3M	B4M	PCT	RND	MAN	CL	SL	B2M	B3M	B4M	PCT	RND	MAN		
BOWSrf	0.458	0.457	0.514	0.506	0.519	0.513	0.472	0.517	0.991	0.991	0.99	0.99	0.993	0.989	0.991	0.537	0.483	
TFIDFSrf	0.48	0.472	0.528	0.53	0.533	0.531	0.489	0.538	0.984	0.989	0.988	0.986	0.99	0.991	0.99	0.557	0.5	
Struc	0.346	0.346	0.384	0.389	0.386	0.392	0.364	0.392	0.991	0.994	0.994	0.993	0.992	0.994	0.991	0.993	0.374	0.292
Pres	0.355	0.349	0.371	0.379	0.373	0.374	0.358	0.371	0.991	0.989	0.99	0.992	0.99	0.991	0.987	0.989	0.354	0.302
Context	0.257	0.253	0.289	0.292	0.284	0.293	0.268	0.283	0.858	0.862	0.861	0.861	0.863	0.861	0.86	0.278	0.275	
PV_R	0.321	0.363	0.405	0.399	0.397	0.389	0.336	0.407	0.996	0.996	0.996	0.996	0.997	0.997	0.996	0.997	0.405	0.323
PV_C	0.375	0.361	0.463	0.467	0.457	0.463	0.402	0.459	0.998	0.997	0.997	0.997	0.998	0.997	0.997	0.999	0.468	0.404
PV_GLR	0.329	0.322	0.388	0.384	0.387	0.383	0.341	0.388	0.997	0.996	0.996	0.997	0.996	0.998	0.997	0.997	0.359	0.264
PV_GLC	0.344	0.342	0.4	0.391	0.39	0.391	0.333	0.392	0.996	0.997	0.997	0.999	0.998	0.998	0.997	0.997	0.394	0.331
Ngrams	0.413	0.417	0.487	0.482	0.481	0.487	0.43	0.493	0.995	0.994	0.995	0.995	0.995	0.995	0.995	0.995	0.487	0.467
<b>SPIRIT</b>																		
BOWSrf	0.621	0.622	0.612	0.626	0.619	0.62	0.62	0.618	0.967	0.969	0.965	0.968	0.969	0.971	0.963	0.966	0.624	0.626
TFIDFSrf	0.622	0.632	0.624	0.623	0.629	0.623	0.627	0.626	0.975	0.975	0.973	0.972	0.975	0.976	0.974	0.97	0.626	0.621
Struc	0.599	0.594	0.607	0.608	0.601	0.601	0.601	0.607	0.959	0.962	0.962	0.961	0.963	0.965	0.965	0.966	0.608	0.598
Pres	0.554	0.549	0.552	0.549	0.554	0.551	0.553	0.554	0.959	0.96	0.953	0.957	0.959	0.958	0.953	0.956	0.55	0.529
Context	0.469	0.472	0.47	0.471	0.467	0.468	0.47	0.474	0.829	0.838	0.825	0.831	0.838	0.837	0.827	0.835	0.46	0.45
PV_R	0.583	0.588	0.588	0.592	0.581	0.58	0.575	0.586	0.978	0.978	0.978	0.979	0.981	0.981	0.976	0.976	0.603	0.585
PV_C	0.573	0.569	0.577	0.578	0.583	0.59	0.574	0.575	0.978	0.981	0.974	0.979	0.977	0.979	0.968	0.975	0.565	0.576
PV_GLR	0.552	0.548	0.555	0.543	0.547	0.558	0.55	0.552	0.983	0.985	0.981	0.985	0.984	0.985	0.982	0.982	0.54	0.545
PV_GLC	0.553	0.552	0.562	0.571	0.555	0.562	0.555	0.563	0.978	0.984	0.976	0.981	0.981	0.98	0.972	0.976	0.556	0.563
Ngrams	0.573	0.576	0.581	0.584	0.58	0.581	0.586	0.979	0.984	0.978	0.979	0.98	0.979	0.978	0.978	0.575	0.586	

Table 10: The performance of the different machine learning tasks applied on the different features using bagging. The hierarchy construction method are abbreviated as follows: balanced  $k$ -means clustering (BkM), predictive clustering tree (PCT), clustering with complete linkage (CL), clustering with single linkage (SL), random (RND) and manual (MAN). The evaluation measure is *macro precision* where larger value means better result.

20-genres	HMC						HSC						MLC	SC				
	CL	SL	B2M	B3M	B4M	PCT	RND	MAN	CL	SL	B2M	B3M	B4M	PCT	RND	MAN		
BOWSrf	0.579	0.641	0.589	0.575	0.588	0.577	0.575	0.584	0.988	0.988	0.987	0.991	0.988	0.988	0.617	0.521		
TFIDFSrf	0.615	0.603	0.584	0.577	0.579	0.572	0.589	0.586	0.99	0.98	0.98	0.993	0.984	0.983	0.981	0.61	0.546	
Struc	0.426	0.442	0.476	0.477	0.463	0.472	0.469	0.471	0.996	0.998	0.997	0.997	0.997	0.997	0.996	0.531	0.333	
Pres	0.423	0.423	0.368	0.362	0.366	0.36	0.373	0.361	0.989	0.99	0.991	0.993	0.991	0.988	0.991	0.378	0.31	
Context	0.277	0.267	0.306	0.307	0.303	0.308	0.286	0.301	0.841	0.848	0.844	0.843	0.842	0.845	0.846	0.31	0.293	
PV_R	0.569	0.561	0.666	0.607	0.664	0.643	0.553	0.669	0.997	0.997	0.998	0.999	0.997	0.998	0.997	0.53	0.348	
PV_C	0.723	0.659	0.623	0.641	0.636	0.634	0.679	0.618	0.999	0.999	0.997	0.997	0.998	0.998	0.999	0.508	0.422	
PV_GI_R	0.595	0.519	0.519	0.52	0.54	0.522	0.555	0.507	0.993	0.998	0.998	0.995	0.998	0.999	0.993	0.533	0.261	
PV_GI_C	0.647	0.552	0.624	0.635	0.635	0.602	0.576	0.61	0.999	0.998	0.999	1	0.999	0.999	0.999	0.477	0.33	
Ngrams	0.595	0.583	0.574	0.586	0.591	0.577	0.607	0.6	0.99	0.989	0.992	0.99	0.991	0.992	0.991	0.592	0.496	
<b>SPIRIT</b>																		
BOWSrf	0.328	0.326	0.304	0.371	0.315	0.329	0.397	0.4	0.99	0.993	0.989	0.991	0.992	0.989	0.992	0.367	0.333	
TFIDFSrf	0.333	0.407	0.31	0.412	0.393	0.375	0.4	0.39	0.925	0.925	0.925	0.997	0.925	0.925	0.924	0.402	0.344	
Struc	0.366	0.326	0.347	0.342	0.336	0.333	0.342	0.337	0.993	0.994	0.994	0.995	0.923	0.99	0.992	0.395	0.355	
Pres	0.175	0.176	0.174	0.171	0.177	0.174	0.173	0.194	0.99	0.991	0.989	0.99	0.918	0.989	0.916	0.254	0.237	
Context	0.238	0.261	0.228	0.239	0.223	0.239	0.258	0.243	0.849	0.79	0.784	0.777	0.783	0.849	0.796	0.787	0.215	0.191
PV_R	0.201	0.203	0.294	0.3	0.29	0.284	0.221	0.297	0.995	0.996	0.996	0.996	0.997	0.997	0.995	0.996	0.42	0.296
PV_C	0.21	0.221	0.222	0.261	0.254	0.248	0.217	0.211	0.996	0.997	0.922	0.925	0.925	0.922	0.925	0.267	0.25	
PV_GI_R	0.211	0.213	0.349	0.254	0.204	0.209	0.207	0.271	0.998	0.998	0.998	0.998	0.998	0.998	0.999	0.378	0.284	
PV_GI_C	0.212	0.218	0.205	0.213	0.2	0.199	0.209	0.21	0.923	0.997	0.923	0.926	0.925	0.922	0.926	0.267	0.231	
Ngrams	0.294	0.3	0.255	0.393	0.287	0.247	0.294	0.238	0.993	0.996	0.992	0.994	0.995	0.996	0.995	0.338	0.339	

Table 11: The performance of the different machine learning tasks applied on the different features using bagging. The hierarchy construction method are abbreviated as follows: balanced  $k$ -means clustering (BkM), predictive clustering tree (PCT), clustering with complete linkage (CL), clustering with single linkage (SL), random (RND) and manual (MAN). The evaluation measure is *macro recall* where larger value means better result.

20-genres	HMC										HSC				MLC		SC		
	CL	SL	B2M	B3M	B4M	PCT	RND	MAN	CL	SL	B2M	B3M	B4M	PCT	RND	MAN			
BOWSrf	0.397	0.396	0.515	0.505	0.511	0.512	0.428	0.51	0.991	0.992	0.989	0.992	0.989	0.992	0.991	0.991	0.527	0.482	
TFIDFSrf	0.434	0.432	0.545	0.547	0.544	0.553	0.458	0.552	0.97	0.996	0.993	0.976	0.995	0.996	0.996	0.997	0.574	0.521	
Struc	0.264	0.263	0.331	0.337	0.329	0.336	0.293	0.339	0.98	0.986	0.989	0.986	0.985	0.987	0.987	0.981	0.989	0.301	0.256
Pres	0.28	0.279	0.332	0.343	0.332	0.332	0.294	0.334	0.989	0.983	0.984	0.99	0.983	0.987	0.987	0.98	0.984	0.298	0.26
Context	0.221	0.214	0.274	0.275	0.269	0.275	0.238	0.268	0.878	0.876	0.881	0.88	0.878	0.879	0.879	0.877	0.261	0.278	
PV_R	0.187	0.319	0.317	0.313	0.31	0.301	0.215	0.323	0.99	0.992	0.993	0.993	0.992	0.991	0.991	0.995	0.395	0.289	
PV_C	0.254	0.242	0.415	0.419	0.392	0.418	0.299	0.416	0.995	0.995	0.995	0.996	0.996	0.995	0.995	0.998	0.512	0.398	
PV_GLR	0.27	0.258	0.39	0.378	0.38	0.382	0.292	0.382	1	0.991	0.991	0.999	0.992	0.994	1	0.993	0.311	0.219	
PV_GLC	0.223	0.224	0.315	0.303	0.301	0.311	0.22	0.311	0.991	0.995	0.994	0.998	0.995	0.995	0.995	0.995	0.394	0.31	
Ngrams	0.35	0.343	0.487	0.482	0.472	0.49	0.372	0.491	0.997	0.998	0.998	0.997	0.997	0.997	0.998	0.471	0.482		
SPIRIT																			
BOWSrf	0.243	0.244	0.223	0.268	0.231	0.232	0.248	0.23	0.795	0.793	0.796	0.818	0.816	0.793	0.795	0.756	0.308	0.3	
TFIDFSrf	0.243	0.249	0.233	0.235	0.252	0.245	0.261	0.249	0.746	0.73	0.719	0.729	0.739	0.715	0.713	0.687	0.308	0.291	
Struc	0.209	0.209	0.219	0.225	0.22	0.223	0.212	0.229	0.697	0.701	0.704	0.71	0.688	0.689	0.851	0.832	0.235	0.245	
Pres	0.195	0.192	0.199	0.195	0.197	0.197	0.195	0.2	0.716	0.699	0.696	0.666	0.701	0.672	0.683	0.668	0.214	0.191	
Context	0.187	0.189	0.192	0.192	0.186	0.188	0.187	0.194	0.539	0.546	0.543	0.54	0.541	0.559	0.539	0.539	0.169	0.168	
PV_R	0.217	0.218	0.193	0.195	0.195	0.193	0.186	0.195	0.831	0.798	0.811	0.812	0.778	0.785	0.788	0.75	0.245	0.238	
PV_C	0.182	0.18	0.189	0.194	0.202	0.206	0.19	0.193	0.794	0.812	0.76	0.792	0.758	0.771	0.764	0.73	0.187	0.202	
PV_GLR	0.18	0.176	0.192	0.179	0.183	0.19	0.187	0.191	0.825	0.842	0.816	0.822	0.812	0.808	0.753	0.181	0.189		
PV_GLC	0.176	0.173	0.188	0.19	0.184	0.189	0.185	0.188	0.8	0.82	0.794	0.813	0.796	0.784	0.776	0.776	0.175	0.188	
Ngrams	0.194	0.197	0.206	0.208	0.216	0.207	0.2	0.21	0.854	0.868	0.877	0.841	0.843	0.834	0.874	0.83	0.215	0.252	

Table 12: The performance of the different machine learning tasks applied on the different features using bagging. The hierarchy construction method are abbreviated as follows: balanced  $k$ -means clustering (BkM), predictive clustering tree (PCT), clustering with complete linkage (CL), clustering with single linkage (SL), random (RND) and manual (MAN). The evaluation measure is *macro F<sub>1</sub>*, where larger value means better result.

20-genres	HMC						HSC						MLC	SC		
	CL	SL	B2M	B3M	B4M	PCT	RND	MAN	CL	SL	B2M	B3M	B4M	PCT	RND	MAN
BOWSrf	0.431	0.433	0.522	0.514	0.523	0.519	0.449	0.519	0.989	0.99	0.988	0.992	0.988	0.99	0.55	0.494
TFIDFSrf	0.459	0.457	0.544	0.544	0.542	0.547	0.48	0.549	0.98	0.988	0.986	0.984	0.989	0.99	0.578	0.531
Struc	0.288	0.291	0.35	0.355	0.346	0.352	0.316	0.357	0.988	0.992	0.993	0.991	0.992	0.989	0.993	0.342
Pres	0.298	0.291	0.328	0.336	0.329	0.329	0.301	0.332	0.989	0.986	0.987	0.991	0.987	0.984	0.987	0.316
Context	0.239	0.231	0.285	0.286	0.281	0.287	0.254	0.279	0.857	0.86	0.858	0.86	0.859	0.859	0.859	0.281
PV_R	0.213	0.307	0.365	0.365	0.358	0.354	0.243	0.377	0.994	0.994	0.995	0.995	0.995	0.996	0.996	0.418
PV_C	0.278	0.261	0.457	0.47	0.44	0.467	0.333	0.458	0.997	0.997	0.996	0.996	0.996	0.997	0.998	0.491
PV_GI_R	0.25	0.233	0.381	0.374	0.376	0.376	0.279	0.375	0.996	0.995	0.995	0.995	0.997	0.997	0.996	0.351
PV_GI_C	0.222	0.227	0.359	0.343	0.346	0.354	0.224	0.355	0.995	0.996	0.996	0.999	0.997	0.997	0.997	0.325
Ngrams	0.384	0.38	0.5	0.5	0.491	0.506	0.4	0.508	0.994	0.993	0.994	0.994	0.993	0.994	0.995	0.311
<hr/>																
SPiRiT																
BOWSrf	0.237	0.234	0.235	0.268	0.245	0.245	0.243	0.245	0.848	0.845	0.857	0.875	0.863	0.848	0.849	0.317
TFIDFSrf	0.236	0.244	0.242	0.251	0.272	0.26	0.258	0.27	0.803	0.785	0.774	0.79	0.791	0.78	0.775	0.306
Struc	0.215	0.208	0.229	0.234	0.228	0.23	0.217	0.238	0.778	0.778	0.787	0.797	0.761	0.764	0.904	0.249
Pres	0.184	0.182	0.184	0.181	0.185	0.183	0.182	0.189	0.79	0.771	0.77	0.73	0.77	0.742	0.759	0.209
Context	0.198	0.2	0.198	0.2	0.191	0.194	0.198	0.199	0.602	0.604	0.606	0.594	0.596	0.618	0.609	0.175
PV_R	0.199	0.2	0.199	0.202	0.199	0.196	0.191	0.204	0.891	0.864	0.88	0.878	0.84	0.854	0.857	0.246
PV_C	0.181	0.181	0.191	0.194	0.201	0.207	0.189	0.191	0.855	0.867	0.816	0.845	0.815	0.826	0.824	0.246
PV_GI_R	0.173	0.169	0.185	0.174	0.176	0.182	0.175	0.183	0.889	0.903	0.885	0.883	0.876	0.881	0.865	0.212
PV_GI_C	0.175	0.173	0.185	0.186	0.181	0.185	0.179	0.186	0.848	0.871	0.844	0.86	0.847	0.836	0.836	0.198
Ngrams	0.197	0.201	0.205	0.212	0.216	0.205	0.199	0.207	0.907	0.915	0.925	0.89	0.893	0.894	0.922	0.231
<hr/>																

Table 13: The performance of the different machine learning tasks applied on the different features using bagging. The hierarchy construction method are abbreviated as follows: balanced  $k$ -means clustering (BkM), predictive clustering tree (PCT), clustering with complete linkage (CL), clustering with single linkage (SL) and manual (MAN). The evaluation measure is *one-error* where smaller value means better result.

20-genres	HMC										HSC				MLC		SC	
	CL	SL	B2M	B3M	B4M	PCT	RND	MAN	CL	SL	B2M	B3M	B4M	PCT	RND	MAN		
BOWSrf	0.496	0.504	0.452	0.461	0.452	0.458	0.494	0.448	0.001	0	0.001	0.001	0.001	0.001	0.002	0.001	0.41	0.438
TFDFSrf	0.487	0.49	0.429	0.431	0.429	0.428	0.468	0.423	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.395	0.417
Struc	0.617	0.614	0.562	0.576	0.583	0.58	0.599	0.575	0.001	0.001	0	0.001	0.001	0	0.001	0	0.591	0.641
Pres	0.615	0.626	0.602	0.6	0.602	0.598	0.602	0.613	0.002	0.002	0.001	0.002	0.002	0.002	0.002	0.001	0.616	0.629
Context	0.723	0.732	0.684	0.678	0.683	0.689	0.712	0.686	0.134	0.131	0.132	0.131	0.135	0.131	0.129	0.679	0.674	
PV_R	0.622	0.631	0.555	0.555	0.573	0.564	0.613	0.565	0	0	0	0	0	0	0	0	0.592	0.634
PV_C	0.577	0.602	0.497	0.499	0.49	0.49	0.548	0.499	0	0	0	0	0	0	0	0	0.496	0.51
PV_GLR	0.658	0.664	0.604	0.599	0.587	0.605	0.65	0.596	0	0	0	0	0	0	0	0	0.609	0.672
PV_GlC	0.617	0.639	0.557	0.567	0.567	0.635	0.567	0	0	0	0	0	0	0	0	0	0.594	0.594
Ngrams	0.555	0.548	0.47	0.473	0.478	0.468	0.524	0.462	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.465	0.438
SPIRIT																		
BOWSrf	0.322	0.308	0.308	0.306	0.304	0.299	0.306	0.015	0.013	0.015	0.013	0.013	0.017	0.016	0.014	0.018	0.308	0.277
TFIDFSrf	0.309	0.302	0.309	0.318	0.312	0.309	0.302	0.009	0.007	0.009	0.01	0.005	0.009	0.011	0.01	0.296	0.298	
Struc	0.321	0.315	0.321	0.321	0.316	0.321	0.316	0.022	0.027	0.026	0.023	0.025	0.021	0.025	0.021	0.306	0.308	
Pres	0.374	0.373	0.374	0.374	0.36	0.37	0.365	0.382	0.02	0.018	0.017	0.018	0.018	0.021	0.022	0.023	0.384	0.382
Context	0.479	0.465	0.477	0.472	0.47	0.461	0.471	0.462	0.13	0.126	0.133	0.133	0.124	0.128	0.133	0.133	0.472	0.481
PV_R	0.35	0.348	0.342	0.336	0.355	0.342	0.342	0.335	0.004	0.004	0.004	0.003	0.003	0.004	0.006	0.005	0.345	0.332
PV_C	0.352	0.359	0.362	0.353	0.359	0.349	0.349	0.36	0.004	0.003	0.001	0.004	0.006	0.001	0.006	0.007	0.346	0.328
PV_GLR	0.394	0.382	0.397	0.389	0.376	0.372	0.384	0.379	0.001	0.002	0.003	0.001	0.004	0.001	0.004	0.002	0.386	0.372
PV_GlC	0.387	0.37	0.37	0.369	0.365	0.373	0.37	0.005	0.003	0.007	0.001	0.003	0.004	0.007	0.005	0.367	0.345	
Ngrams	0.365	0.349	0.36	0.357	0.355	0.363	0.357	0.007	0.004	0.006	0.003	0.006	0.006	0.006	0.005	0.359	0.34	

Table 14: The performance of the different machine learning tasks applied on the different features using bagging. The hierarchy construction method are abbreviated as follows: balanced  $k$ -means clustering (BkM), predictive clustering tree (PCT), clustering with complete linkage (CL), clustering with single linkage (SL), random (RND) and manual (MAN). The evaluation measure is coverage where smaller value means better result.

20-genres	HMC						HSC						MLC	SC		
	CL	SL	B2M	B3M	B4M	PCT	RND	MAN	CL	SL	B2M	B3M	B4M	PCT	RND	MAN
BOWSrf	3.143	3.06	2.62	2.715	2.703	2.642	3.155	2.645	0.342	0.342	0.342	0.343	0.343	0.344	2.484	2.748
TFIDFSrf	2.887	2.929	2.474	2.513	2.452	2.474	2.903	2.394	0.342	0.342	0.342	0.342	0.342	0.342	2.404	2.602
Struc	4.122	4.175	3.8	3.767	3.841	3.771	4.128	3.882	0.342	0.342	0.342	0.342	0.342	0.342	3.763	5.017
Pres	4.827	4.838	4.511	4.65	4.554	4.525	4.65	4.622	0.344	0.343	0.344	0.343	0.344	0.344	4.582	4.809
Context	6.094	6.157	5.675	5.791	5.746	5.743	5.919	5.756	0.713	0.711	0.708	0.709	0.704	0.711	0.707	5.936
PV_R	3.947	4.018	3.544	3.432	3.516	3.471	3.749	3.339	0.342	0.342	0.342	0.342	0.342	0.342	3.451	4.32
PV_C	3.467	3.501	2.77	2.776	2.827	2.73	3.18	2.803	0.342	0.342	0.342	0.342	0.342	0.342	2.865	3.424
PV_GI_R	4.621	4.562	3.82	3.799	3.882	3.857	4.38	3.87	0.342	0.342	0.342	0.342	0.342	0.342	0.342	5.176
PV_GI_C	4.032	4.123	3.434	3.428	3.385	3.38	3.902	3.495	0.342	0.342	0.342	0.342	0.342	0.342	3.589	3.979
Ngrams	3.275	3.262	2.808	2.829	2.85	2.807	3.253	2.775	0.343	0.343	0.343	0.343	0.343	0.343	2.81	2.77
<b>SPIRIT</b>																
BOWSrf	1.963	1.955	1.899	1.902	1.915	1.936	1.956	1.902	0.591	0.589	0.591	0.592	0.594	0.59	0.593	1.906
TFIDFSrf	2	2.014	1.904	1.965	1.959	1.97	1.93	0.584	0.582	0.584	0.585	0.585	0.582	0.585	0.584	1.996
Struc	2.06	2.105	2.048	2.084	2.104	2.109	2.101	2.077	0.6	0.604	0.604	0.601	0.599	0.6	0.597	2.026
Pres	2.377	2.35	2.284	2.312	2.291	2.285	2.311	2.333	0.601	0.596	0.599	0.601	0.599	0.601	0.604	2.366
Context	2.732	2.725	2.708	2.704	2.725	2.681	2.753	2.73	0.837	0.832	0.838	0.839	0.828	0.833	0.833	2.526
PV_R	2.231	2.197	2.133	2.16	2.196	2.184	2.218	2.203	0.579	0.578	0.578	0.577	0.577	0.577	0.579	3.071
PV_C	2.166	2.172	2.142	2.136	2.098	2.166	2.14	0.58	0.577	0.577	0.578	0.577	0.577	0.578	0.579	2.349
PV_GI_R	2.384	2.34	2.298	2.339	2.323	2.305	2.384	2.298	0.576	0.576	0.576	0.577	0.577	0.577	0.575	2.241
PV_GI_C	2.227	2.18	2.173	2.15	2.138	2.132	2.22	2.172	0.581	0.578	0.582	0.577	0.579	0.578	0.584	2.468
Ngrams	2.203	2.18	2.162	2.199	2.157	2.142	2.209	2.172	0.584	0.578	0.584	0.579	0.58	0.581	0.58	2.201
																2.333

Table 15: The performance of the different machine learning tasks applied on the different features using bagging. The hierarchy construction method are abbreviated as follows: balanced  $k$ -means clustering (BkM), predictive clustering tree (PCT), clustering with complete linkage (CL), clustering with single linkage (SL), random (RND) and manual (MAN). The evaluation measure is *ranking loss* where smaller value means better result.

20-genres	HMC										HSC			MLC		SC	
	CL	SL	B2M	B3M	B4M	PCT	RND	MAN	CL	SL	B2M	B3M	B4M	PCT	RND	MAN	
BOWSrf	0.123	0.119	0.098	0.103	0.102	0.099	0.122	0.099	0	0	0	0	0	0	0	0.092	0.102
TFIDFSrf	0.111	0.114	0.092	0.094	0.09	0.092	0.112	0.088	0	0	0	0	0	0	0	0.088	0.097
Struc	0.168	0.171	0.154	0.153	0.156	0.154	0.168	0.158	0	0	0	0	0	0	0	0.154	0.212
Pres	0.206	0.207	0.189	0.195	0.191	0.19	0.197	0.195	0	0	0	0	0	0	0	0.194	0.202
Context	0.269	0.271	0.247	0.251	0.25	0.261	0.251	0.018	0.018	0.018	0.018	0.018	0.018	0.018	0.018	0.26	0.261
PV_R	0.164	0.166	0.146	0.14	0.145	0.142	0.156	0.136	0	0	0	0	0	0	0	0.143	0.18
PV_C	0.143	0.144	0.107	0.107	0.109	0.106	0.127	0.109	0	0	0	0	0	0	0	0.111	0.135
PV_GLR	0.198	0.196	0.159	0.157	0.162	0.16	0.185	0.162	0	0	0	0	0	0	0	0.164	0.223
PV_GLC	0.168	0.172	0.136	0.137	0.137	0.136	0.161	0.139	0	0	0	0	0	0	0	0.145	0.162
Ngrams	0.13	0.129	0.106	0.108	0.109	0.107	0.129	0.105	0	0	0	0	0	0	0	0.107	0.104
SPIRIT																	
BOWSrf	0.084	0.083	0.079	0.079	0.079	0.08	0.082	0.079	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.078	0.078
TFIDFSrf	0.085	0.086	0.079	0.082	0.081	0.08	0.083	0.08	0.001	0.001	0.001	0	0.001	0.001	0.001	0.08	0.082
Struc	0.089	0.092	0.088	0.09	0.092	0.092	0.09	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.086	0.101
Pres	0.111	0.109	0.105	0.107	0.106	0.106	0.108	0.108	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.109	0.118
Context	0.136	0.136	0.135	0.134	0.136	0.132	0.138	0.136	0.017	0.017	0.017	0.017	0.017	0.017	0.017	0.147	0.153
PV_R	0.101	0.1	0.096	0.097	0.099	0.098	0.099	0.099	0	0	0	0	0	0	0	0.098	0.105
PV_C	0.097	0.098	0.096	0.096	0.096	0.092	0.097	0.095	0	0	0	0	0.001	0	0.001	0.096	0.1
PV_GLR	0.11	0.108	0.106	0.109	0.105	0.111	0.106	0	0	0	0	0	0	0	0	0.106	0.115
PV_GLC	0.101	0.098	0.099	0.096	0.097	0.096	0.102	0.098	0	0	0.001	0	0	0	0.001	0.099	0.104
Ngrams	0.1	0.098	0.097	0.098	0.097	0.096	0.099	0.098	0.001	0	0.001	0.001	0.001	0	0.001	0.097	0.094

Table 16: The performance of the different machine learning tasks applied on the different features using bagging. The hierarchy construction method are abbreviated as follows: balanced  $k$ -means clustering (BkM), predictive clustering tree (PCT), clustering with complete linkage (CL), clustering with single linkage (SL), random (RND) and manual (MAN). The evaluation measure is *average precision* where larger value means better result.

20-genres	HMC						HSC						MLC		SC	
	CL	SL	B2M	B3M	B4M	PCT	RND	MAN	CL	SL	B2M	B3M	B4M	PCT	RND	
BOWSrf	0.629	0.627	0.668	0.659	0.666	0.665	0.632	0.669	1	1	1	1	1	1	0.999	0.999
TFIDFSrf	0.641	0.635	0.683	0.679	0.684	0.681	0.648	0.687	1	1	1	1	1	1	1	0.7
Struc	0.528	0.529	0.566	0.561	0.554	0.562	0.537	0.559	1	1	1	1	1	1	1	0.554
Pres	0.511	0.504	0.526	0.525	0.528	0.52	0.518	0.999	0.999	0.999	0.999	0.999	0.999	0.999	0.999	0.517
Context	0.417	0.411	0.449	0.449	0.445	0.445	0.429	0.445	0.914	0.915	0.916	0.915	0.916	0.914	0.916	0.444
PV_R	0.527	0.523	0.581	0.585	0.575	0.577	0.542	0.582	1	1	1	1	1	1	1	0.568
PV_C	0.568	0.553	0.634	0.637	0.634	0.64	0.595	0.633	1	1	1	1	1	1	1	0.634
PV_GI_R	0.485	0.483	0.544	0.547	0.552	0.546	0.5	0.546	1	1	1	1	1	1	1	0.536
PV_GI_C	0.523	0.512	0.578	0.576	0.57	0.574	0.522	0.573	1	1	1	1	1	1	1	0.547
Ngrams	0.59	0.596	0.647	0.644	0.641	0.646	0.604	0.653	1	0.999	0.999	0.999	0.999	0.999	0.999	0.65
<b>SPIRIT</b>																
BOWSrf	0.765	0.768	0.774	0.773	0.772	0.769	0.771	0.771	0.992	0.993	0.992	0.991	0.992	0.991	0.991	0.771
TFIDFSrf	0.765	0.769	0.772	0.769	0.771	0.77	0.773	0.996	0.996	0.995	0.997	0.996	0.995	0.995	0.995	0.777
Struc	0.751	0.753	0.755	0.753	0.755	0.754	0.752	0.756	0.988	0.986	0.988	0.988	0.988	0.988	0.987	0.751
Pres	0.714	0.716	0.718	0.715	0.719	0.715	0.719	0.71	0.989	0.991	0.99	0.99	0.989	0.988	0.988	0.702
Context	0.644	0.65	0.646	0.649	0.648	0.655	0.648	0.651	0.921	0.923	0.92	0.924	0.923	0.92	0.921	0.634
PV_R	0.733	0.733	0.738	0.74	0.731	0.738	0.735	0.741	0.998	0.998	0.998	0.998	0.998	0.997	0.997	0.738
PV_C	0.734	0.73	0.731	0.733	0.734	0.739	0.735	0.734	0.998	0.999	0.999	0.998	0.999	0.996	0.996	0.732
PV_GI_R	0.702	0.713	0.707	0.71	0.713	0.718	0.706	0.714	0.999	0.999	0.999	1	0.998	0.998	0.999	0.712
PV_GI_C	0.718	0.725	0.728	0.729	0.731	0.722	0.726	0.997	0.998	0.996	0.999	0.998	0.996	0.997	0.724	0.732
Ngrams	0.731	0.738	0.736	0.736	0.735	0.736	0.73	0.734	0.998	0.996	0.996	0.997	0.997	0.997	0.997	0.748