

# タスク説明&評価結果報告

2018/10/18

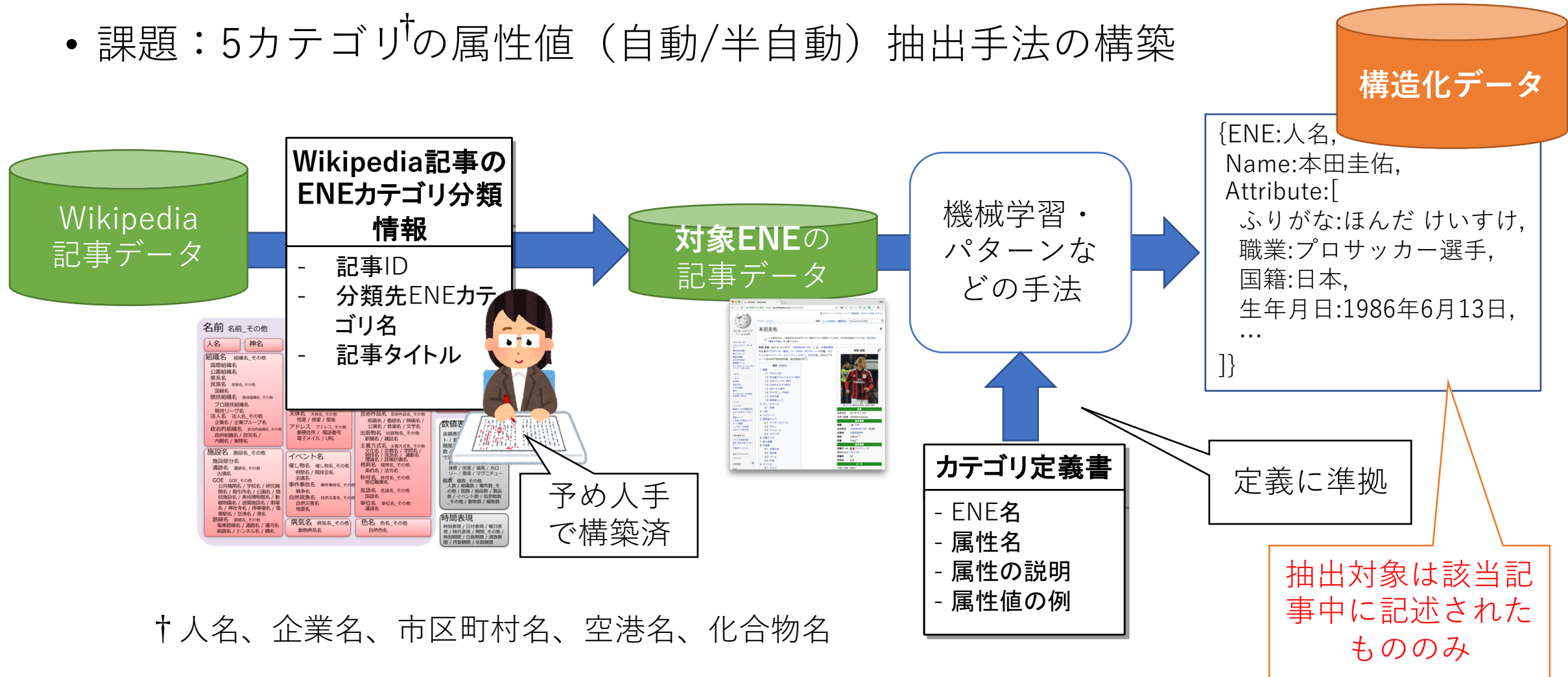
森羅プロジェクト2018実行委員 副委員長 小林暁雄

# 参加チーム

- 全参加チーム（7チーム+1チーム）
  - 大学：
    - 豊橋技術科学大学（応用数理ネットワーク研究室）様
    - 岡山県立大学大学院知能情報処理研究室 石井颯人様
    - 長岡技術科学大学 野中尋史様
  - 企業
    - 富士ゼロックス 杉原大悟様
    - Sansan株式会社様
    - 凸版印刷株式会社様
    - 日本ユニシス株式会社様
  - AIP
    - 中山功太（インターン）

# 構造化タスク概要

- 課題：5カテゴリ<sup>†</sup>の属性値（自動/半自動）抽出手法の構築



# カテゴリとチーム手法一覧

	人名	企業名	市区町村名	空港名	化合物名
豊技大	Pattern	Pattern	Pattern	Pattern	
岡山県立大	Pattern + Heuristic				
長技大					ML(LightGBM) + Pattern
Sansan		Pattern			
富士ゼロックス	DL(NCRFpp)	DL(NCRFpp)	DL(NCRFpp)	DL(NCRFpp)	DL(NCRFpp)
				DL(NCRFpp)	DL(RDFDNN)
		DL(RDFDNN)	DL(RDFDNN)	DL(RDFDNN)	DL(NCRFpp)
				DL(RDFDNN)	DL(RDFDNN)
凸版印刷		Pattern(B)		DL(A)	
		DL(A)			
		Pattern(B+C)			
		DL(A+B+C)			
日本ユニシス	DL(DRQA)	DL(DRQA)	DL(DRQA)	DL(DRQA)	DL(DRQA)
中山 (AIP)	DL	DL	DL	DL	DL

青字：参考手法、DL（深層学習ベース）、ML（その他機械学習）

# 全体の結果(マイクロF値)

チーム	手法	人名	企業名	市区町村名	空港名	化合物名
豊技大		<b>0.20</b>	<b>0.41</b>	<b>0.28</b>	<u><b>0.72</b></u>	
岡山県立大	Pattern(参考)	0.19				
	Pattern + Heuristic(参考)	0.16				
長技大						0.42
Sansan			<b>0.30</b>			
富士ゼロックス	NCRFpp	0.31			<b>0.38</b>	<b>0.15</b>
	RDFDNN					<b>0.15</b>
	NCRFpp(参考)		0.30	0.43	0.42	0.39
	RDFDNN(参考)		0.28		0.40	0.37
凸版印刷	A		<b>0.29</b>		<b>0.35</b>	
	B		<b>0.33</b>			
	B+C (参考)		0.34		0.42	
	A+B+C (参考)		0.41			
日本ユニシス	DRQA		<u><b>0.53</b></u>		<b>0.67</b>	<u><b>0.47</b></u>
	DRQA(参考)	<u>0.44</u>		0.42		
中山 (AIP)	(参考)	0.36	0.38	<u>0.46</u>	0.71	0.46

# 人名の結果

	0001:豊橋技術科学大学 (応用数理)			0011:日本ユニシス (参考)			0012:杉原大悟 (参考)			0016:中山功太 (AIP) (参考)			岡山県立大学大学院知能情報処理研究			岡山県立大学大学院知能情報処理研究		
attribute	precision	recall	f1-score	precision	recall	f1-score	precision	recall	f1-score	precision	recall	f1-score	precision	recall	f1-score	precision	recall	f1-score
ふりがな	<u>0.958</u>	0.315	0.474	0.905	0.781	<u>0.838</u>	0.028	0.027	0.027	0.437	<u>0.945</u>	0.598	0.387	0.329	0.356	0.387	0.329	0.356
両親	0.286	0.043	0.075	0.85	0.37	0.516	0.75	0.13	0.222	0.239	0.348	0.283	0.122	0.13	0.126	0.122	0.13	0.126
代表作	0.25	0.023	0.042	0.225	0.307	0.26	0	0	0	0.036	0.114	0.055	0	0	0	0	0	0
別名	0.22	0.094	0.132	0.61	0.18	0.278	0.33	0.245	<u>0.281</u>	0.192	<u>0.475</u>	0.273	0.556	0.144	0.229	2	0.144	0.229
参加イベント	<u>1</u>	0.005	0.01	0.285	0.312	<u>0.298</u>	0.433	0.153	0.226	0.206	<u>0.413</u>	0.275	0	0	0	0	0	0
受賞歴	0.167	0.01	0.019	0.438	0.189	0.264	<u>0.505</u>	0.186	0.272	0.412	<u>0.385</u>	<u>0.398</u>	0.015	0.003	0.005	0.015	0.003	0.005
国籍	<u>1</u>	<u>0.538</u>	<u>0.7</u>	0.971	0.872	<u>0.919</u>	<u>0.538</u>	<u>0.897</u>	<u>0.673</u>	0.197	<u>0.974</u>	0.328	<u>0.314</u>	<u>0.821</u>	<u>0.454</u>	<u>0.314</u>	<u>0.821</u>	<u>0.454</u>
学歴	0.529	0.061	0.109	<u>0.544</u>	0.5	0.521	0.534	0.264	0.353	0.398	<u>0.689</u>	0.505	0.824	0.095	0.17	<u>0.824</u>	0.095	0.17
家族	0.432	0.075	0.128	<u>0.76</u>	0.178	0.288	0.358	0.136	0.197	0.223	<u>0.477</u>	<u>0.304</u>	0.41	0.192	0.262	0.41	0.192	0.262
居住地	0.651	0.262	0.374	0.503	0.463	<u>0.482</u>	0.392	0.341	0.365	0.243	<u>0.724</u>	0.364	<u>0.762</u>	0.075	0.137	<u>0.762</u>	0.075	0.137
師匠	0	0	0	<u>0.5</u>	<u>0.167</u>	<u>0.25</u>	0	0	0	0.022	0.056	0.032	0	0	0	0	0	0
所属組織	0.175	0.058	0.087	<u>0.543</u>	0.268	<u>0.359</u>	0.449	0.291	0.353	0.257	<u>0.563</u>	0.353	0.508	0.06	0.107	0.508	0.06	0.107
時代	<u>0.75</u>	0.231	0.353	0.429	0.231	0.3	0.556	0.385	<u>0.455</u>	0.375	<u>0.462</u>	0.414	0	0	0	0	0	0
本名	<u>0.52</u>	<u>0.191</u>	<u>0.279</u>	<u>0.661</u>	<u>0.574</u>	<u>0.614</u>	0.311	0.279	0.294	0.453	0.353	0.397	0.409	0.397	0.403	0.409	0.397	0.403
死因	<u>1</u>	0.375	0.545	0.773	<u>0.708</u>	<u>0.739</u>	0.412	0.292	0.342	0.238	0.417	0.303	0	0	0	0	0	0
没地	0.5	0.217	0.303	<u>0.529</u>	<u>0.391</u>	<u>0.45</u>	0	0	0	0	0	0	0	0	0	0	0	0
没年月日	0.077	0.029	0.042	0.929	<u>0.743</u>	<u>0.826</u>	0.462	0.686	0.552	<u>0.957</u>	0.629	0.759	0	0	0	0	0	0
生年月日	0.848	0.626	0.72	<u>0.944</u>	0.794	<u>0.863</u>	0.582	0.729	0.647	0.659	<u>0.85</u>	0.742	0.92	0.757	0.831	0.462	0.056	0.1
生誕地	0.348	0.419	0.38	<u>0.769</u>	0.538	<u>0.633</u>	0	0	0	0.225	<u>0.667</u>	0.336	0.2	0.011	0.021	0.2	0.011	0.021
称号	0.235	0.093	0.133	0.229	<u>0.256</u>	0.242	<u>0.5</u>	0.186	<u>0.271</u>	0.093	0.163	0.118	0	0	0	0	0	0
職業	0.511	0.11	0.181	0.689	0.297	<u>0.415</u>	0.393	0.256	0.31	0.329	<u>0.543</u>	0.41	<u>0.765</u>	0.179	0.29	<u>0.765</u>	0.179	0.29
異表記	0	0	0	0.732	<u>0.833</u>	<u>0.779</u>	<u>1</u>	0.139	0.244	0.5	0.694	0.581	0.033	0.306	0.06	0.033	0.306	0.06

※赤字:提出結果で最も良いスコア 青字:全体で最も良いスコア (参考) アンダーバー:全体で最も良いスコア (提出結果)

■: Infoboxに出現しやすい (出現頻度0.7以上) ■: 記事冒頭に出現しやすい (0.7以上)

■: Infoboxと記事冒頭両方に出現しやすい

# 企業名の結果

	0006: 凸版印刷株式会社 B			0005: 凸版印刷株式会社 A			0001: 豊橋技術科学大学 (応用数理ネットワー			0004: Sansan株式会社			0007: 日本ユニシス		
attribute	TOPPAN_B pr	TOPPAN_B re	TOPPAN_B f1	TOPPAN_A pr	TOPPAN_A re	TOPPAN_A f1	TUT precisior	TUT recall	TUT f1-score	SANSAN prec	SANSAN reca	SANSAN f1-s	UNISYS precisi	UNISYS recall	UNISYS f1-sco
正式名称	0	0	0	0.459	0.53	0.492	0.955	0.75	0.84	0.943	0.66	0.777	0.744	0.465	0.572
ふりがな	0.451	0.787	0.573	0.75	0.83	0.788	0.041	0.085	0.055	0	0	0	0.974	0.809	0.884
別名	0.288	0.508	0.368	0.079	0.169	0.108	0.938	0.508	0.659	0	0	0	0.686	0.407	0.511
種類	0.988	0.919	0.952	0.872	0.872	0.872	1	0.93	0.964	0.986	0.814	0.892	0.899	0.93	0.914
本拠地国	0	0	0	0.71	0.917	0.8	0.986	0.75	0.852	0.968	0.635	0.767	0.952	0.833	0.889
本拠地	0.58	0.379	0.458	0.459	0.294	0.358	0.667	0.34	0.45	0.475	0.307	0.373	0.706	0.392	0.504
設立年	0.8	0.5	0.615	0.333	0.319	0.326	0.983	0.362	0.529	0.455	0.281	0.347	1	0.544	0.705
業界	0.765	0.886	0.821	0.19	0.091	0.123	0.933	0.477	0.631	0.551	0.432	0.484	0.938	0.852	0.893
事業内容	0.71	0.143	0.238	0.293	0.205	0.241	0.553	0.199	0.293	0.708	0.149	0.246	0.74	0.275	0.401
取扱商品	0	0	0	0.322	0.329	0.325	0	0	0	0	0	0	0.413	0.065	0.112
代表者	0.821	0.621	0.707	0.12	0.029	0.047	0.596	0.544	0.569	0.354	0.272	0.308	0.924	0.709	0.802
資本金	0.985	0.89	0.935	0	0	0	1	0.822	0.902	0.667	0.575	0.618	0.984	0.863	0.92
資本金データの	1	0.241	0.388	0.053	0.034	0.041	0.75	0.517	0.612	0.286	0.138	0.186	1	0.759	0.863
従業員数 (単体)	0.949	0.755	0.841	0	0	0	0.818	0.735	0.774	0.783	0.367	0.5	0.935	0.878	0.906
従業員数 (単体)	0.864	0.487	0.623	0	0	0	0.194	0.179	0.186	0.269	0.179	0.215	0.838	0.795	0.816
従業員数 (連結)	0.929	0.867	0.897	0	0	0	0.933	0.933	0.933	0	0	0	0.933	0.933	0.933
従業員数 (連結)	0.833	0.312	0.454	0	0	0	0.333	0.312	0.322	0	0	0	0.833	0.625	0.714
売上高 (単体)	0.875	0.212	0.341	0	0	0	1	0.848	0.918	0	0	0	1	0.909	0.952
売上高データの	0	0	0	0	0	0	0.885	0.742	0.807	0	0	0	1	0.903	0.949
売上高 (連結)	1	0.158	0.273	0	0	0	0.471	0.842	0.604	1	0.053	0.101	0.941	0.842	0.889
売上高 (連結)	0	0	0	0	0	0	0.469	0.789	0.588	1	0.579	0.733	1	0.895	0.945
主要株主	0.632	0.155	0.249	0.161	0.239	0.192	0.427	0.226	0.296	0.224	0.097	0.135	0.329	0.148	0.204
子会社・合併会	0.18	0.071	0.102	0.156	0.233	0.187	0.667	0.119	0.202	0.771	0.146	0.246	0.46	0.091	0.152
業界内地位・規	0	0	0	0.094	0.103	0.098	0	0	0	0	0	0	0.75	0.207	0.324
買収・合併した	0	0	0	0.115	0.114	0.114	0	0	0	0	0	0	0.36	0.146	0.208
起源	0	0	0	0.106	0.25	0.149	0	0	0	1	0.05	0.095	0.6	0.3	0.4
過去の社名	0	0	0	0.137	0.188	0.158	0	0	0	0	0	0	0.584	0.542	0.562
創業国	0	0	0	0.504	0.903	0.647	0	0	0	0	0	0	0.726	0.847	0.782
創業地	0	0	0	0.197	0.406	0.265	0	0	0	0	0	0	0.9	0.562	0.692
創業者	0	0	0	0.122	0.214	0.155	0	0	0	0	0	0	0.538	0.5	0.518
創業時の事業	0	0	0	0.103	0.121	0.111	0	0	0	0	0	0	0.444	0.242	0.313
社名使用開始年	0	0	0	0.233	0.247	0.24	0	0	0	0	0	0	0.377	0.358	0.367
コーポレートス	0	0	0	0	0	0	0	0	0	0	0	0	0.75	0.462	0.572

# 企業名の結果 (参考)

	0012: 杉原大悟 (富士ゼロックス) (参考)			0013: 杉原大悟 (富士ゼロックス) rdfnn (参			0014: 凸版印刷株式会社 B + C (参考)			0015: 凸版印刷株式会社 A + B + C (参考)			0016: 中山功太 (AIP) (参考)		
attribute	precision	recall	f1-score	precision	recall	f1-score	precision	recall	f1-score	precision	recall	f1-score	precision	recall	f1-score
正式名称	0.303	0.28	0.291	0.303	0.28	0.291	0	0	0	0.459	0.53	0.492	0.674	0.445	0.536
ふりがな	0.83	0.83	0.83	0.83	0.83	0.83	0.425	0.787	0.552	0.413	0.915	0.569	0.627	0.894	0.737
別名	0.091	0.051	0.065	0.091	0.051	0.065	0.288	0.508	0.368	0.159	0.61	0.252	0.657	0.39	0.489
種類	0.857	0.977	0.913	0.857	0.977	0.913	0.988	0.919	0.952	0.875	0.977	0.923	0.802	0.942	0.866
本拠地国	0.798	0.74	0.768	0.798	0.74	0.768	0	0	0	0.71	0.917	0.8	0.923	0.75	0.828
本拠地	0.75	0.098	0.173	0.047	0.105	0.065	0.613	0.444	0.515	0.51	0.647	0.57	0.769	0.458	0.574
設立年	0.765	0.406	0.53	0.765	0.406	0.53	0.8	0.5	0.615	0.49	0.731	0.587	0.939	0.388	0.549
業界	0.735	0.284	0.41	0.735	0.284	0.41	0.765	0.886	0.821	0.581	0.898	0.706	0.855	0.602	0.707
事業内容	0.58	0.19	0.286	0.58	0.19	0.286	0.71	0.143	0.238	0.386	0.345	0.364	0.616	0.342	0.44
取扱商品	0.481	0.186	0.268	0.398	0.186	0.254	0	0	0	0.322	0.329	0.325	0.23	0.418	0.297
代表者	0.049	0.039	0.043	0.049	0.039	0.043	0.821	0.621	0.707	0.65	0.65	0.65	0.897	0.505	0.646
資本金	0.652	0.589	0.619	0.652	0.589	0.619	0.985	0.89	0.935	0.833	0.89	0.861	0.875	0.575	0.694
資本金データの	0	0	0	0	0	0	1	0.241	0.388	0.308	0.276	0.291	0.5	0.345	0.408
従業員数 (単体)	0.196	0.204	0.2	0.196	0.204	0.2	0.949	0.755	0.841	0.627	0.755	0.685	0.688	0.449	0.543
従業員数 (単体)	0	0	0	0	0	0	0.864	0.487	0.623	0.679	0.487	0.567	0.477	0.538	0.506
従業員数 (連結)	0	0	0	0	0	0	0.929	0.867	0.897	0.929	0.867	0.897	1	0.2	0.333
従業員数 (連結)	0	0	0	0	0	0	0.833	0.312	0.454	0.075	0.312	0.121	0.714	0.625	0.667
売上高 (単体)	0.05	0.061	0.055	0.05	0.061	0.055	0.875	0.212	0.341	0.778	0.212	0.333	0.905	0.576	0.704
売上高データの	0	0	0	0	0	0	0	0	0	0	0	0	0.615	0.516	0.561
売上高 (連結)	0	0	0	0	0	0	1	0.158	0.273	1	0.158	0.273	0.875	0.737	0.8
売上高 (連結)	0	0	0	0	0	0	0	0	0	0	0	0	0.371	0.684	0.481
主要株主	0.165	0.09	0.116	0.165	0.09	0.116	0.632	0.155	0.249	0.21	0.355	0.264	0.714	0.226	0.343
子会社・合併会	0.241	0.051	0.084	0.241	0.051	0.084	0.236	0.083	0.123	0.189	0.304	0.233	0.143	0.316	0.197
業界内地位・規	0	0	0	0	0	0	0	0	0	0.086	0.103	0.094	0.077	0.103	0.088
買収・合併した	0.406	0.106	0.168	0.406	0.106	0.168	0	0	0	0.115	0.114	0.114	0.114	0.268	0.16
起源	0.5	0.3	0.375	0.5	0.3	0.375	0	0	0	0.106	0.25	0.149	0.097	0.3	0.147
過去の社名	0.564	0.323	0.411	0.564	0.323	0.411	0	0	0	0.137	0.188	0.158	0.175	0.51	0.261
創業国	0	0	0	0	0	0	0	0	0	0.504	0.903	0.647	0.688	0.764	0.724
創業地	0.625	0.156	0.25	0.625	0.156	0.25	0	0	0	0.197	0.406	0.265	0.124	0.406	0.19
創業者	0.25	0.179	0.209	0.25	0.179	0.209	0	0	0	0.122	0.214	0.155	0.09	0.464	0.151
創業時の事業	0.538	0.212	0.304	0.538	0.212	0.304	0	0	0	0.103	0.121	0.111	0.07	0.182	0.101
社名使用開始年	0	0	0	0	0	0	0	0	0	0.233	0.247	0.24	0.125	0.593	0.206
コーポレートス	0	0	0	0	0	0	1	0.154	0.267	0.105	0.154	0.125	0.034	0.077	0.047



# 市区町村名の結果

attribute	0001:豊橋技術科学大学 (応用数理ネ)			0011:日本ユニシス (参考)			0012:杉原大悟 (富士ゼロックス)			0016:中山功太 (AIP) (参考)		
	precision	recall	f1-score	precision	recall	f1-score	precision	recall	f1-score	precision	recall	f1-score
ふりがな	0.957	0.918	0.937	0.953	0.837	0.891	0.936	0.898	0.917	0.935	0.878	0.906
人口	0.551	0.392	0.458	0.831	0.66	0.736	0.588	0.412	0.485	0.667	0.454	0.54
人口密度	1	0.92	0.958	0.917	0.44	0.595	0.778	0.42	0.546	0.95	0.38	0.543
人口データの年	0.69	0.444	0.54	0.955	0.711	0.815	0.667	0.089	0.157	0.787	0.656	0.716
別名	0.447	0.298	0.358	0.68	0.362	0.472	0.532	0.411	0.464	0.335	0.638	0.439
友好市区町村	0.054	0.728	0.101	0.809	0.259	0.392	0.893	0.68	0.772	0.502	0.741	0.599
合併市区町村	0	0	0	0.107	0.023	0.038	0.411	0.298	0.345	0.173	0.489	0.256
国	0.934	0.723	0.815	0.926	0.642	0.758	0.671	0.832	0.743	0.865	0.839	0.852
国内位置	0.206	0.438	0.28	0.778	0.438	0.56	0.5	0.062	0.11	0.173	0.562	0.265
地名の由来	0	0	0	0.333	0.056	0.096	0	0	0	0	0	0
座標・経度	0.88	0.886	0.883	0.963	0.394	0.559	0	0	0	0.87	0.811	0.839
座標・緯度	0.895	0.895	0.895	0.426	0.173	0.246	1	0.008	0.016	0.84	0.789	0.814
恒例行事	0	0	0	0.37	0.241	0.292	0.719	0.277	0.4	0.319	0.349	0.333
成立年	0.474	0.108	0.176	0.442	0.277	0.341	0.684	0.157	0.255	0.159	0.723	0.261
所在地	0.511	0.226	0.313	0.753	0.288	0.417	0.635	0.533	0.58	0.758	0.613	0.678
旧称	0.025	0.058	0.035	0.318	0.269	0.291	0.083	0.019	0.031	0.063	0.231	0.099
温泉・鉱泉	0.4	0.364	0.381	0.4	0.182	0.25	0	0	0	0.571	0.364	0.445
特産品	0	0	0	0.389	0.286	0.33	0.341	0.306	0.323	0.203	0.633	0.307
産業	0	0	0	0.375	0.188	0.25	0.477	0.637	0.546	0.23	0.7	0.346
種類	0.316	0.563	0.405	0.776	0.598	0.675	0.439	0.701	0.54	0.349	0.782	0.483
著名な地形	0.429	0.231	0.3	0	0	0	0	0	0	0.048	0.077	0.059
著名施設	0.103	0.034	0.051	0	0	0	0.286	0.023	0.043	0.043	0.091	0.058
観光地	0.064	0.008	0.014	0.464	0.035	0.065	0.382	0.092	0.148	0.255	0.478	0.333
鉄道会社	0	0	0	0.517	0.484	0.5	0.565	0.203	0.299	0.46	0.719	0.561
面積	0.789	0.155	0.259	0.833	0.258	0.394	0.771	0.278	0.409	0.764	0.433	0.553
首長	0.786	0.268	0.4	0.793	0.561	0.657	0.368	0.171	0.233	0.812	0.634	0.712

# 空港名の結果

	0005: 凸版印刷株式会社 A			0002: 杉原大悟 (富士ゼロックス)			0001: 豊橋技術科学大学 (応用数理)			0007: 日本ユニシス			0012: 杉原大悟 (富士ゼロックス)			0013: 杉原大悟 (富士ゼロックス)			0016: 中山功太 (AIP) (参考)		
attribute	precision	recall	f1-score	precision	recall	f1-score	precision	recall	f1-score	precision	recall	f1-score	precision	recall	f1-score	precision	recall	f1-score	precision	recall	f1-score
ふりがな	0.693	0.825	0.753	0.872	0.54	0.667	0	0	0	1	0.825	0.904	0.862	0.794	0.827	0.862	0.794	0.827	0.73	0.857	0.788
IATA (空港コード)	0.846	0.117	0.206	0.8	0.213	0.336	1	1	1	0.816	0.755	0.784	0.92	0.245	0.387	0.92	0.245	0.387	0.869	0.989	0.925
ICAO (空港コード)	0.556	0.102	0.172	0.706	0.122	0.208	1	0.98	0.99	0.969	0.969	0.969	0.895	0.173	0.29	0.895	0.173	0.29	0.989	0.959	0.974
別名	0.617	0.388	0.476	0.57	0.367	0.447	0.736	0.376	0.498	0.917	0.633	0.749	0.559	0.424	0.482	0.559	0.424	0.482	0.375	0.457	0.412
名称由来	0.118	0.375	0.18	0.667	0.125	0.211	0	0	0	0.429	0.375	0.4	0.667	0.125	0.211	0.022	0.125	0.037	0.8	0.25	0.381
名称由来人物の地位職業名	0.273	0.333	0.3	1	0.111	0.2	0	0	0	0.467	0.389	0.424	1	0.056	0.106	0.108	0.222	0.145	0.312	0.278	0.294
国	0.742	0.871	0.801	0.866	0.741	0.799	0.97	0.691	0.807	0.99	0.705	0.824	0.813	0.813	0.813	0.813	0.813	0.813	0.91	0.95	0.93
年間利用客数	0.583	0.311	0.406	0.725	0.644	0.682	0.933	0.622	0.746	0.891	0.911	0.901	0.844	0.6	0.701	0.844	0.6	0.701	0.812	0.578	0.675
年間利用者数データの年	0.5	0.087	0.148	0.778	0.152	0.254	0.829	0.63	0.716	0.929	0.848	0.887	0.667	0.174	0.276	0.667	0.174	0.276	0.684	0.565	0.619
年間発着回数	0	0	0	0.25	0.034	0.06	0.92	0.793	0.852	0.929	0.897	0.913	0	0	0	0	0	0	0.875	0.724	0.792
年間発着回数データの年	0	0	0	0	0	0	0.686	0.857	0.762	0.962	0.893	0.926	0	0	0	0	0	0	0.786	0.786	0.786
座標・経度	0	0	0	0	0	0	0.585	0.985	0.734	1	0.495	0.662	0	0	0	0	0	0	0.995	0.974	0.984
座標・緯度	0	0	0	0	0	0	0.583	0.985	0.732	0	0	0	0	0	0	0	0	0	0.995	0.974	0.984
所在地	0.585	0.504	0.541	0.758	0.472	0.582	0.847	0.337	0.482	0.875	0.455	0.599	0.712	0.533	0.61	0.712	0.533	0.61	0.742	0.667	0.703
旧称	0.171	0.212	0.189	0.471	0.242	0.32	0	0	0	0.409	0.545	0.467	0.545	0.182	0.273	0.545	0.182	0.273	0.19	0.364	0.25
標高	0.091	0.006	0.011	0.266	0.123	0.168	0.994	0.947	0.97	0.967	0.52	0.676	0.266	0.123	0.168	0.266	0.123	0.168	1	0.936	0.967
母都市	0.2	0.92	0.329	0.688	0.44	0.537	0.882	0.6	0.714	0.941	0.64	0.762	0.591	0.52	0.553	0.591	0.52	0.553	0.483	0.56	0.519
滑走路数	0.692	1	0.818	0.75	0.667	0.706	0	0	0	0.75	1	0.857	0.818	1	0.9	0.818	1	0.9	0.375	1	0.545
滑走路の長さ	0.864	0.114	0.201	1	0.12	0.214	0.986	0.832	0.902	0.968	0.551	0.702	1	0.12	0.214	1	0.12	0.214	0.315	0.138	0.192
総面積	0	0	0	1	0.6	0.75	0	0	0	0.7	0.7	0.7	0.857	0.6	0.706	0.857	0.6	0.706	0.167	0.3	0.215
近隣空港	0.098	0.172	0.125	0.167	0.034	0.056	0	0	0	0.355	0.379	0.367	0.077	0.034	0.047	0.077	0.034	0.047	0.075	0.138	0.097
運営者	0.276	0.18	0.218	0.371	0.596	0.457	0.897	0.685	0.777	0.87	0.674	0.76	0.377	0.618	0.468	0.377	0.618	0.468	0.737	0.629	0.679
運用時間	0.071	0.071	0.071	0.167	0.143	0.154	1	0.857	0.923	1	0.857	0.923	0.167	0.143	0.154	0.167	0.143	0.154	1	0.786	0.88
開港年	0.56	0.509	0.533	0.88	0.4	0.55	1	0.218	0.358	0.821	0.582	0.681	0.857	0.545	0.666	0.857	0.545	0.666	0.342	0.745	0.469

※赤字:提出結果で最も良いスコア 青字:全体で最も良いスコア (参考) アンダーバー:全体で最も良いスコア (提出結果)

■: Infoboxに出現しやすい (出現頻度0.7以上) ■: 記事冒頭に出現しやすい (0.7以上)

■: Infoboxと記事冒頭両方に出現しやすい

# 化合物名の結果

	0009: 野中尋史 (長岡技術科学大学)			0002: 杉原大悟 (富士ゼロックス)			0003: 杉原大悟 (富士ゼロックス)			0007: 日本ユニシス			0012: 杉原大悟 (富士ゼロックス)			0013: 杉原大悟 (富士ゼロックス)			0016: 中山功太 (AIP) (参考)		
attribute	precision	recall	f1-score	precision	recall	f1-score	precision	recall	f1-score	precision	recall	f1-score	precision	recall	f1-score	precision	recall	f1-score	precision	recall	f1-score
ふりがな	0	0	0	0	0	0	0	0	0	<u>0.971</u>	<u>0.917</u>	<u>0.943</u>	0.725	0.806	0.763	0.725	0.806	0.763	0.718	0.778	0.747
別称	0.829	<u>0.641</u>	<u>0.723</u>	0.4	0.006	0.012	0.4	0.006	0.012	<u>0.838</u>	0.441	0.578	0.562	0.291	0.383	0.562	0.291	0.383	0.561	0.553	0.557
用途	0.284	<u>0.184</u>	0.223	0	0	0	0.017	0.005	0.008	<u>0.611</u>	0.167	<u>0.262</u>	<u>0.641</u>	0.298	<u>0.407</u>	0.396	0.303	0.343	0.341	<u>0.48</u>	0.399
種類	0.514	0.444	0.476	0	0	0	0	0	0	<u>0.82</u>	<u>0.651</u>	<u>0.726</u>	<u>0.821</u>	0.508	0.628	<u>0.821</u>	0.508	0.628	0.692	0.643	0.667
商標名	0	0	0	0	0	0	0	0	0	<u>0.769</u>	<u>0.333</u>	<u>0.465</u>	0	0	0	0	0	0	0.538	<u>0.7</u>	<u>0.608</u>
特性	0.223	0.12	0.156	<u>0.797</u>	0.192	0.309	0.371	<u>0.247</u>	0.297	0.626	0.234	<u>0.341</u>	0.709	0.396	<u>0.508</u>	0.456	<u>0.422</u>	0.438	0.734	0.341	0.466
原材料	0.246	0.209	0.226	0	0	0	0	0	0	<u>0.495</u>	<u>0.218</u>	<u>0.303</u>	0.478	0.267	<u>0.343</u>	0.478	0.267	<u>0.343</u>	0.239	<u>0.568</u>	0.336
製造方法	0.122	0.038	0.058	0	0	0	0	0	0	<u>0.426</u>	<u>0.22</u>	<u>0.29</u>	0	0	0	0	0	0	0.154	0.136	0.144
生成化合物	0	0	0	0	0	0	0	0	0	<u>0.571</u>	<u>0.183</u>	<u>0.277</u>	0.171	0.046	0.072	0.171	0.046	0.072	0.118	0.107	0.112
CAS番号	<u>1</u>	<u>1</u>	<u>1</u>	0.525	0.323	0.4	0.525	0.323	0.4	<u>1</u>	0.697	0.821	0.438	0.323	0.372	0.438	0.323	0.372	0.989	0.919	0.953
化学式	0.968	<u>0.827</u>	<u>0.892</u>	0.984	0.564	0.717	0.984	0.564	0.717	<u>0.987</u>	0.682	0.807	0.931	0.736	0.822	0.931	0.736	0.822	0.929	0.827	0.875
密度	0.804	<u>0.837</u>	<u>0.82</u>	0.583	0.429	0.494	0.583	0.429	0.494	<u>1</u>	0.429	0.6	0.583	0.429	0.494	0.583	0.429	0.494	0.842	0.653	0.736
融点	0.679	<u>0.812</u>	<u>0.74</u>	0.326	0.125	0.181	0.326	0.125	0.181	<u>0.833</u>	0.268	0.406	0.311	0.125	0.178	0.311	0.125	0.178	0.565	0.116	0.192
沸点	0.897	<u>0.821</u>	<u>0.857</u>	0.143	0.053	0.077	0.143	0.053	0.077	<u>0.963</u>	0.274	0.427	0.133	0.063	0.086	0.133	0.063	0.086	0.483	0.147	0.225
示性式	<u>1</u>	<u>0.818</u>	<u>0.9</u>	0	0	0	0	0	0	<u>1</u>	0.727	0.842	0	0	0	0	0	0	0.667	0.727	0.696

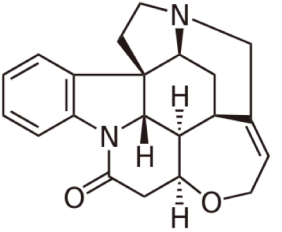
※赤字:提出結果で最も良いスコア 青字:全体で最も良いスコア (参考) アンダーバー:全体で最も良いスコア (提出結果)

 : Infoboxに出現しやすい (出現頻度0.7以上)  : 記事冒頭に出現しやすい (0.7以上)

 : Infoboxと記事冒頭両方に出現しやすい

# ENEカテゴリとInfobox

**(-)ストリキニーネ**



**IUPAC名** [\[表示\]](#)

**別称** [\[表示\]](#)

**識別情報**

CAS登録番号 [57-24-9](#)

PubChem [441971](#)

**複雑なテンプレート**

- ・ Chembox(化合物名)
- ・ 基礎情報\_会社 (企業名)

サブテンプレートが多く、パターンの網羅が困難と考えられる。Infoboxにも本文にも出現する属性値が多い


**Sフレーズ** [S1/2](#) [S36/37](#) [S45](#) [S60](#) [S61](#)

**半数致死量** 2.35 mg/kg (ラット、経口)

**LD50** [\[表示\]](#)

**出典**

**TOYOTA**



トヨタ自動車本社

種類株式会社

市場情報東証1部 **7203**

大証1部 (廃止) **7203**

名証1部 **7203**

福証 **7203**

札証 **7203**

NYSE TM

略称トヨタ / TMC

本社所在地 日本

〒471-8571 愛知県豊田市トヨタ町1番地

本店所在地 〒\*\*\*\*

〇〇県〇〇市〇〇

設立1937年8月28日

業種輸送用機器

**単一のテンプレート**

- ・ Infobox\_空港 (空港名)

多くの属性値がInfoboxから抽出可能

単独：6万8240人 (2014年3月31日現在)

支店舗数〇〇店 (〇〇年〇〇月現在)

決算期毎年3月31日

主要株主日本トラスティ・サービス信託銀行 9.61%

豊田自動織機 6.48%

日本スタートラスト信託銀行 5.27%

**成田国際空港**  
Narita International Airport




**IATA: NRT - ICAO: RJAA**

**概要**

国・地域 ● 日本

所在地 千葉県成田市

母都市 東京

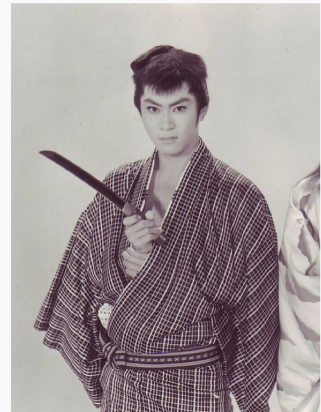
運用者 成田国際空港株式会社

運用開始 23 (14)

座標 ● 北緯35度45分50秒 東経140度23分30秒



きたおおじ きんや  
**北大路 欣也**



映画『若君と次男坊』（東映・1961年）のスクリーン写真より

本名	浅井 隆
生年月日	1943
出生地	<span style="color:red">●</span>
身長	174 cm
血液型	A型
職業	俳優
ジャンル	映画・
活動期間	1956

● 日本の政治家

**田中 角栄**  
たなか かくえい



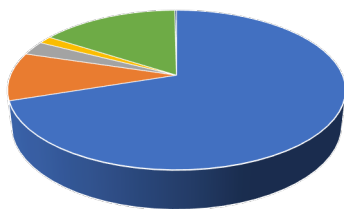
**多様なテンプレート**  
(人名、市区町村名)

共通する属性が少なくなるため、パターンの網羅が困難。記事によってInfoboxと本文どちらに値が出現しやすいかが異なる

四角名	輪島 大士
本名	輪島 博
愛称	
生年月日	1948年1月1日
没年月日	2014年11月1日
出身	新潟県新井市
身長	172 cm
体重	70 kg
BMI	38.57
所属部屋	花籠部屋
得意技	左四つ 寄り 下手投げ
成績	(自由党→) (自由民主党→) 無所属 陸軍上等兵 西山町名誉町民 一級建築士 田中はな 辻和子(妻)

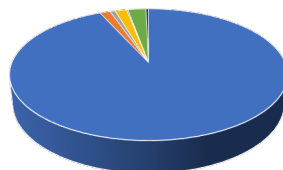
# Infoboxの比率(トレーニングデータ)

化合物名



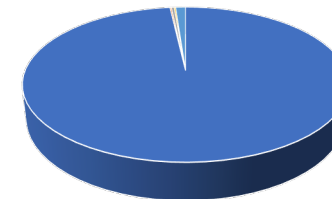
- Chembox
- Infobox\_有機化合物
- Infobox\_無機化合物
- Infobox\_化合物
- Infobox\_爆薬
- Drugbox
- Infobox\_gene

企業名



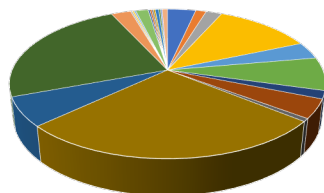
- 基礎情報\_会社
- 基礎情報\_銀行
- 日本の上テ兼営局
- Infobox\_publisher2
- Infobox\_broadcasting\_network
- Infobox\_Company
- 航空会社情報ボックス

空港名



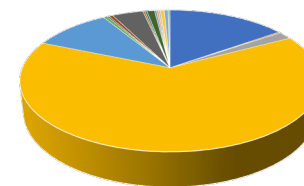
- Infobox\_空港
- Infobox\_JGSDF\_camp
- Infobox\_airport
- Infobox\_Military\_Structure
- 基礎情報\_会社

人名



- 基礎情報\_武士
- 基礎情報\_軍人
- Infobox\_人物
- Infobox\_baseball\_player
- Infobox\_作家
- Infobox\_Musician
- Infobox\_漫画家
- Infobox\_プロレスラー
- Infobox\_お笑い芸人
- ActorActress
- 政治家
- サッカー選手
- バスケットボール選手
- 基礎情報\_アナウンサー
- Infobox\_軍人
- 基礎情報\_天皇
- Infobox\_Officeholder
- Infobox\_陸上選手
- Infobox\_建築家
- Infobox\_motorcycle\_rider
- Infobox\_共和政ローマ
- Infobox\_力士
- Infobox\_scientist
- Infobox\_Engineer

市区町村名



- コムーネ
- 基礎情報\_中国の都市
- Infobox\_UK\_place
- 世界の市
- 基礎情報\_過去の中国の都市
- Infobox\_Settlement
- 基礎情報\_過去の中国の都市
- 基礎情報\_スペインの自治体2
- 基礎情報\_台湾行政区
- 基礎情報\_ベトナムの省
- 基礎情報\_イスラエルの自治体
- Infobox\_Australian\_place
- Infobox\_country
- Infobox\_国

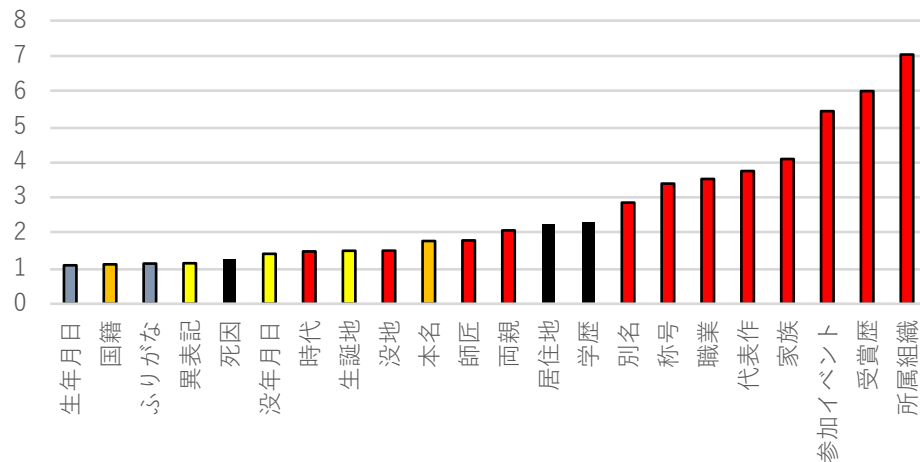
# 次回に向けての結果分析

- Ensemble Learning によって高品質 & 大規模なKBが構築できるか
    - 全結果の総和(union)によって正解の属性値が網羅できるかチェック
      - 手法を組み合わせることであまりいきそうな属性
        - Collaborative Contribution
      - 全ての手法であまり行かなかった属性
        - 次回以降のタスクの重要な対象に
    - 正解データの見直し※で修正が多かった属性に対する調査
      - 人手で見落としやすい属性に強い手法の発見
- ※提出結果の中に正解データ中で見落とししていた属性値がないか改めてチェックを行った

# 全結果の総和から見た 人名

- Infoboxに出現しない属性値の抽出再現率が低い傾向
- Infoboxから抽出しやすい属性が少なく、再現率が低い属性多数。
- Infobox、記事冒頭文から抽出しやすい属性は再現率高

属性値の頻度



attribute	precision	recall	f1-score	best_recall
ふりがな	0.375	0.945	0.537	0.945
両親	0.302	0.565	0.394	0.37
代表作	0.091	0.42	0.15	0.307
別名	0.189	0.64	0.292	0.475
参加イベント	0.175	0.513	0.261	0.413
受賞歴	0.35	0.49	0.408	0.385
国籍	0.194	1	0.325	0.974
学歴	0.351	0.858	0.498	0.689
家族	0.225	0.57	0.323	0.477
居住地	0.219	0.813	0.345	0.724
師匠	0.052	0.167	0.079	0.167
所属組織	0.22	0.641	0.328	0.563
時代	0.304	0.538	0.388	0.462
本名	0.397	0.735	0.516	0.574
死因	0.333	0.75	0.461	0.708
没地	0.139	0.435	0.211	0.391
没年月日	0.478	0.943	0.634	0.743
生年月日	0.483	0.944	0.639	0.85
生誕地	0.211	0.796	0.334	0.667
称号	0.142	0.442	0.215	0.256
職業	0.28	0.591	0.38	0.543
異表記	0.471	0.917	0.622	0.833

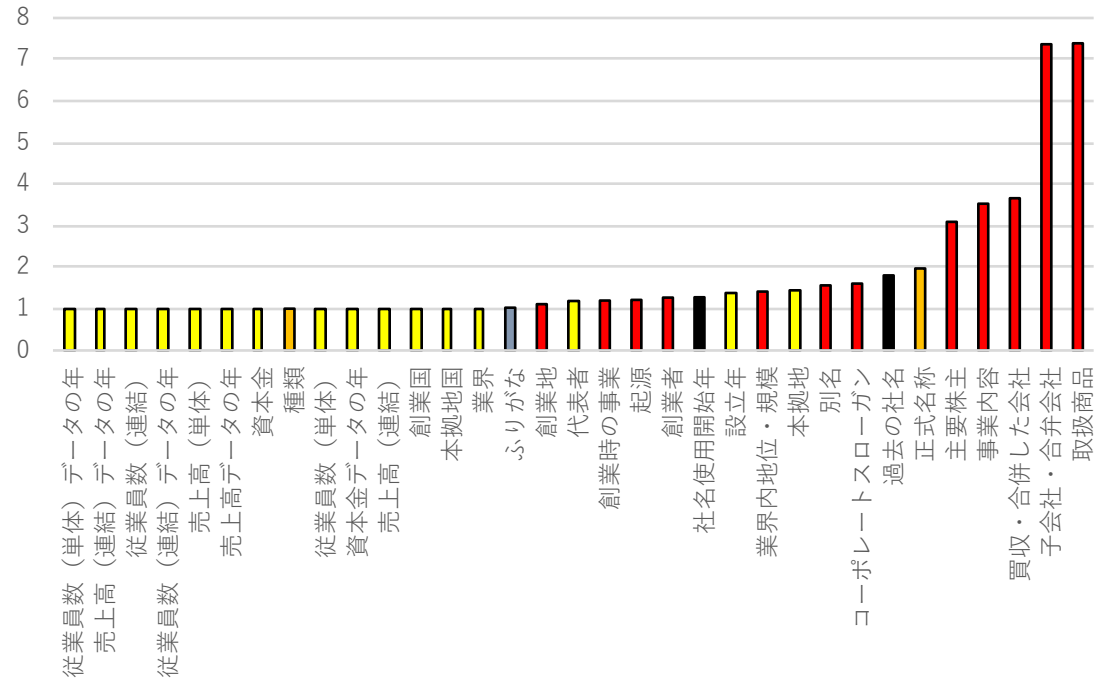
赤字：recall 0.7未満

# 結果の総和から見た企業名

Infoboxに出現しない属性値の抽出再現率が低い傾向

- ・特に、値を**列挙**するタイプの属性（**子会社・合併会社** など）
- ・**取扱商品**や**事業内容**のような**フレーズ**を含む属性も再現率低

属性値頻度



attribute	precision	recall	f1-score	best_recall
正式名称	0.369	0.955	0.532	<b>0.75</b>
ふりがな	0.205	0.979	0.339	0.915
別名	0.15	<b>0.695</b>	0.247	<b>0.61</b>
種類	0.723	1	0.839	0.977
本拠地	0.648	0.979	0.78	<b>0.917</b>
本拠地	0.185	0.752	0.297	0.647
設立年	0.41	0.838	0.551	0.731
業界	0.439	0.932	0.597	0.898
事業内容	0.375	<b>0.596</b>	0.46	0.345
取扱商品	0.211	<b>0.551</b>	0.305	0.418
代表者	0.335	0.883	0.486	<b>0.709</b>
資本金	0.576	0.932	0.712	<b>0.89</b>
資本金データの年	0.373	0.862	0.521	<b>0.759</b>
従業員数 (単体)	0.368	1	0.538	<b>0.878</b>
従業員数 (単体) データの年	0.316	0.949	0.474	<b>0.795</b>
従業員数 (連結)	0.938	1	0.968	<b>0.933</b>
従業員数 (連結) データの年	0.161	0.938	0.275	<b>0.625</b>
売上高 (単体)	0.44	1	0.611	<b>0.909</b>
売上高データの年	0.556	0.968	0.706	<b>0.903</b>
売上高 (連結)	0.487	1	0.655	<b>0.842</b>
売上高 (連結) データの年	0.302	1	0.464	<b>0.895</b>
主要株主	0.159	<b>0.516</b>	0.243	<b>0.239</b>
子会社・合併会社	0.134	<b>0.565</b>	0.217	0.316
業界内地位・規模	0.115	<b>0.31</b>	0.168	<b>0.207</b>
買収・合併した会社	0.129	<b>0.48</b>	0.203	0.268
起源	0.104	<b>0.6</b>	0.177	<b>0.3</b>
過去の社名	0.17	0.781	0.279	<b>0.542</b>
創業国	0.493	0.958	0.651	<b>0.903</b>
創業地	0.127	<b>0.625</b>	0.211	<b>0.562</b>
創業者	0.08	<b>0.607</b>	0.141	<b>0.5</b>
創業時の事業	0.125	<b>0.515</b>	0.201	<b>0.242</b>
社名使用開始年	0.13	0.778	0.223	0.593
コーポレートスローガン	0.145	<b>0.615</b>	0.235	<b>0.462</b>

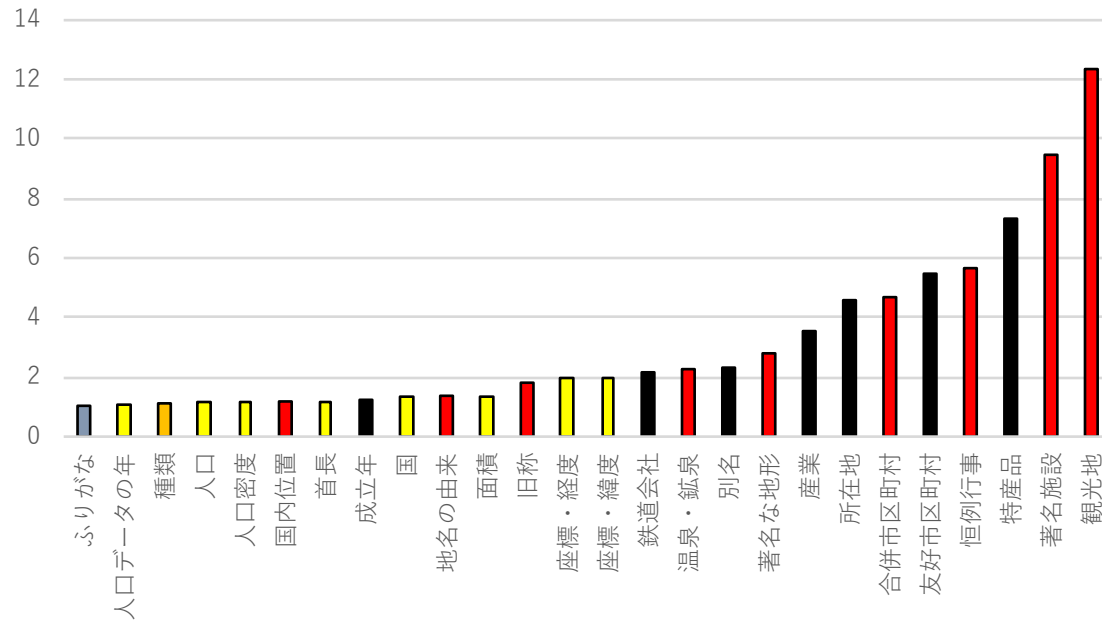


# 結果の総和から見た 市区町村名

人名・企業名と同様の傾向

ただし、国内位置や旧称などの列挙型ではない属性の再現率低  
・記述箇所が特定しづらい、あるいはフレーズ型の属性

属性値の頻度



attribute	precision	recall	f1-score	best_recall
ふりがな	0.836	0.939	0.885	0.918
人口	0.528	0.866	0.656	0.66
人口密度	0.836	0.92	0.876	0.92
人口データの年	0.622	0.933	0.746	0.711
別名	0.265	0.794	0.397	0.638
友好市区町村	0.063	0.918	0.118	0.741
合併市区町村	0.154	0.573	0.243	0.489
国	0.577	0.905	0.705	0.839
国内位置	0.122	0.562	0.2	0.562
地名の由来	0.333	0.056	0.096	0.056
座標・経度	0.826	0.902	0.862	0.886
座標・緯度	0.688	0.977	0.807	0.895
恒例行事	0.318	0.59	0.413	0.349
成立年	0.162	0.819	0.27	0.723
所在地	0.513	0.769	0.615	0.613
旧称	0.074	0.481	0.128	0.269
温泉・鉱泉	0.353	0.545	0.428	0.364
特産品	0.187	0.735	0.298	0.633
産業	0.237	0.787	0.364	0.7
種類	0.279	0.954	0.432	0.782
著名な地形	0.148	0.308	0.2	0.231
著名施設	0.058	0.148	0.083	0.091
観光地	0.242	0.519	0.33	0.478
鉄道会社	0.32	0.766	0.451	0.719
面積	0.722	0.722	0.722	0.433
首長	0.585	0.927	0.717	0.634

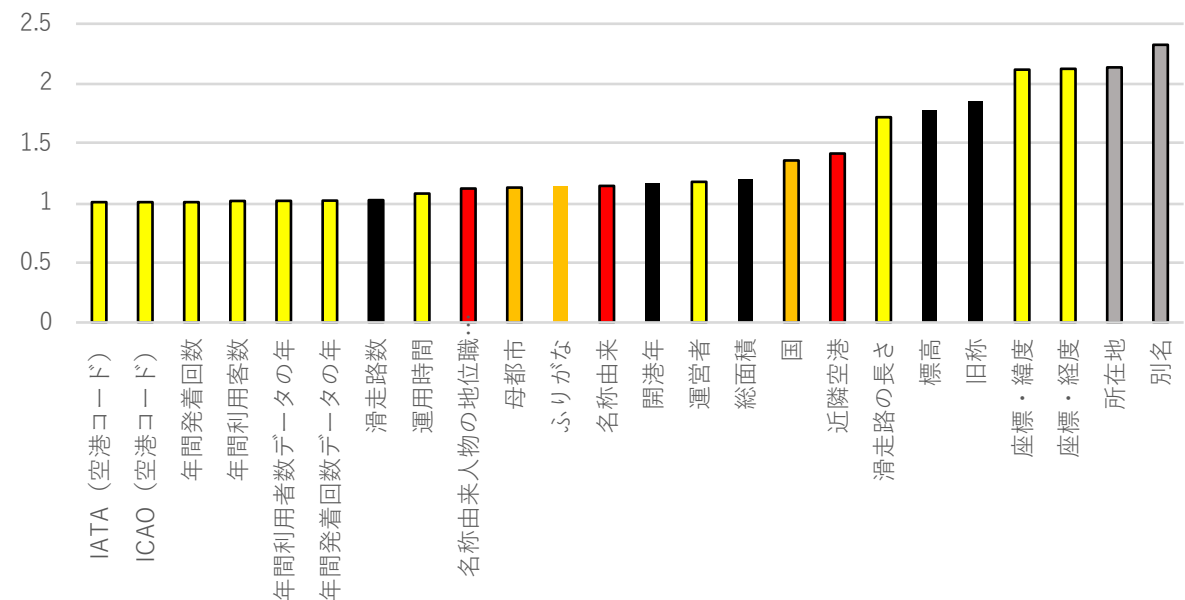
# 全結果の総和からみた空港名

attribute	precision	recall	f1-score	best_recall
ふりがな	0.56	0.968	0.71	0.857
IATA (空港コード)	0.734	1	0.847	1
ICAO (空港コード)	0.858	0.99	0.919	0.98
別名	0.377	0.898	0.531	0.633
名称由来	0.072	0.688	0.13	0.375
名称由来人物の地位職業名	0.137	0.556	0.22	0.389
国	0.66	0.993	0.793	0.95
年間利用客数	0.606	0.956	0.742	0.911
年間利用者数データの年	0.618	0.913	0.737	0.848
年間発着回数	0.667	0.897	0.765	0.897
年間発着回数データの年	0.481	0.929	0.634	0.893
座標・経度	0.587	0.995	0.738	0.985
座標・緯度	0.584	0.995	0.736	0.985
所在地	0.509	0.841	0.634	0.667
旧称	0.186	0.788	0.301	0.545
標高	0.695	0.959	0.806	0.947
母都市	0.197	1	0.329	0.92
滑走路数	0.36	1	0.529	1
滑走路の長さ	0.741	0.976	0.842	0.832
総面積	0.29	0.9	0.439	0.7
近隣空港	0.107	0.448	0.173	0.379
運営者	0.318	0.876	0.467	0.685
運用時間	0.353	0.857	0.5	0.857
開港年	0.307	0.836	0.449	0.745

企業名～市区町村名とは異なり、列挙型・フレーズ型でも再現率高  
多くの属性がInfoboxから抽出可能

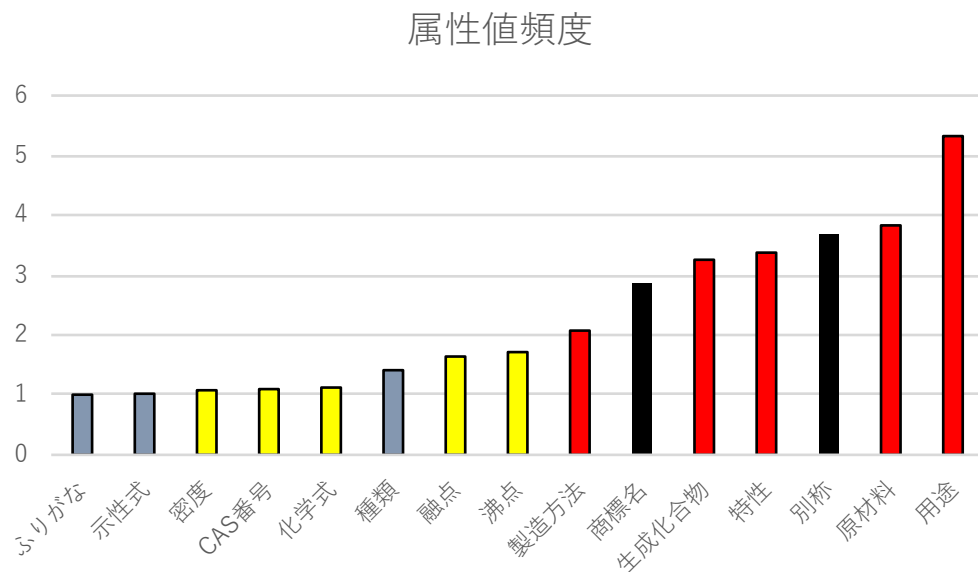
- ・ Infobox以外から抽出する必要のある属性の一部の再現率低
- ・ 名称由来はフレーズ型

属性出現頻度



# 結果の総和から見た化合物名

- ・ 空港名以外と同様の傾向
- ・ **別称**などの列挙型の属性でも、記述箇所が特定しやすい属性は再現率高



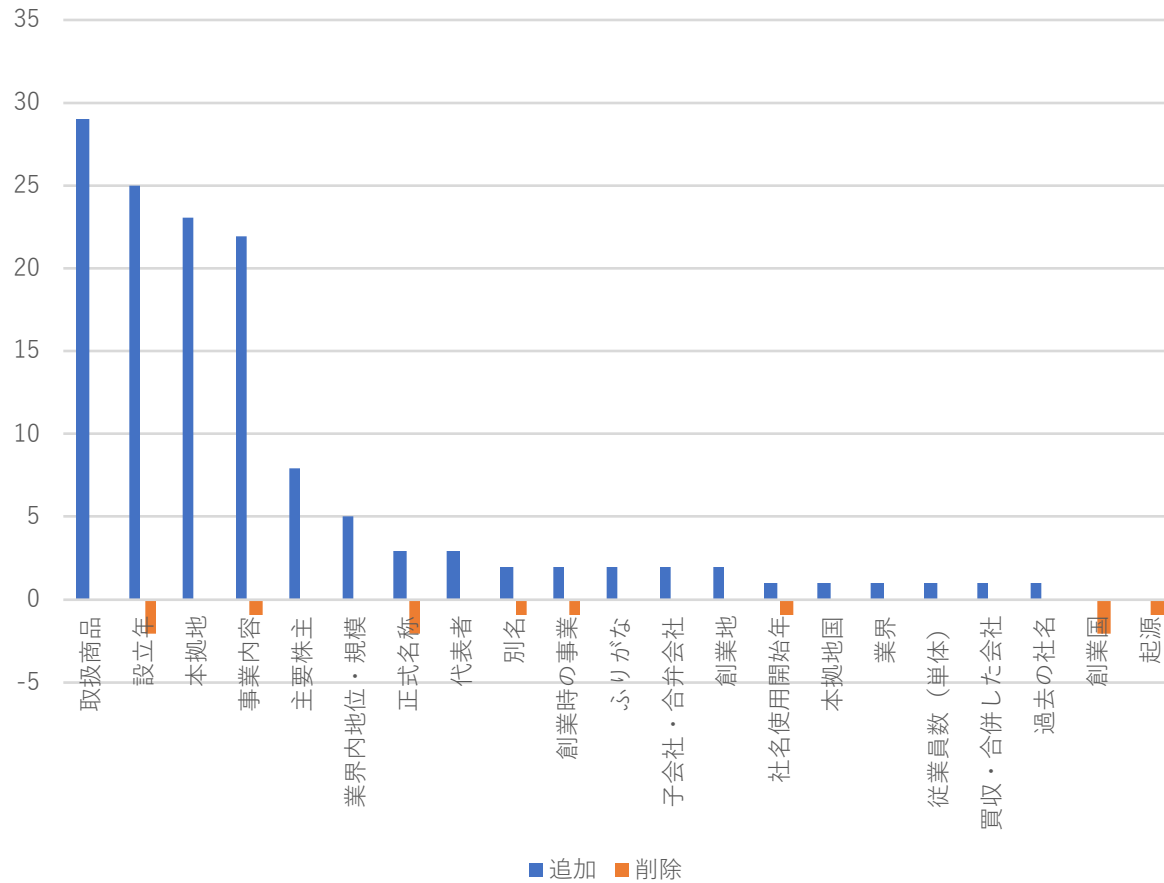
attribute	precision	recall	f1-score	best_recall
ふりがな	0.1	0.972	0.181	<u>0.917</u>
別称	0.502	0.879	0.639	<u>0.641</u>
用途	0.261	0.629	0.369	0.48
種類	0.489	0.865	0.625	<u>0.651</u>
商標名	0.5	0.8	0.615	0.7
特性	0.381	0.63	0.475	0.422
原材料	0.205	0.684	0.315	0.568
製造方法	0.217	0.356	0.27	<u>0.22</u>
生成化合物	0.162	0.206	0.181	<u>0.183</u>
CAS番号	0.702	1	0.825	1
化学式	0.864	0.927	0.894	<u>0.827</u>
密度	0.592	0.918	0.72	<u>0.837</u>
融点	0.519	0.866	0.649	<u>0.812</u>
沸点	0.556	0.842	0.67	<u>0.821</u>
示性式	0.455	0.909	0.606	<u>0.818</u>

# Collabulative Contribution的に

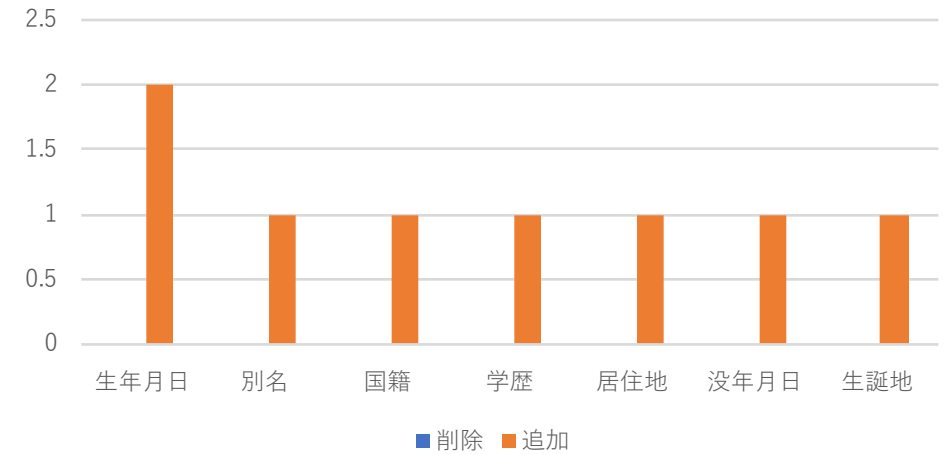
- 本文から抽出する属性&列挙数が多い場合、再現率が低い傾向
- いくつかの属性について、最高性能の手法<<総和
  - Infobox or 本文 や 列挙型/フレーズ型 などとは無関係？  
→今後調査
- トレーニング・テストデータ作成に今回の結果は有効
  - 実際に修正を行った結果を次ページ以降で紹介

# 正解データの修正結果(1/2)

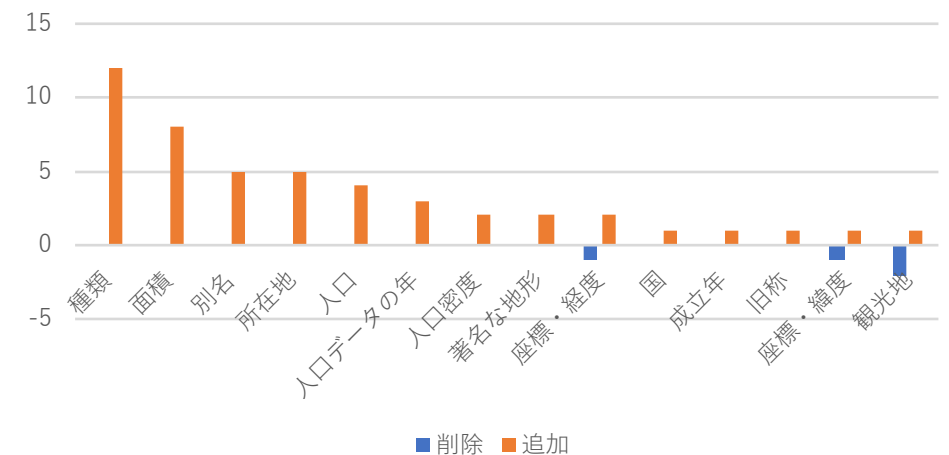
修正された属性値数 (企業名)



修正された属性値数 (人名)

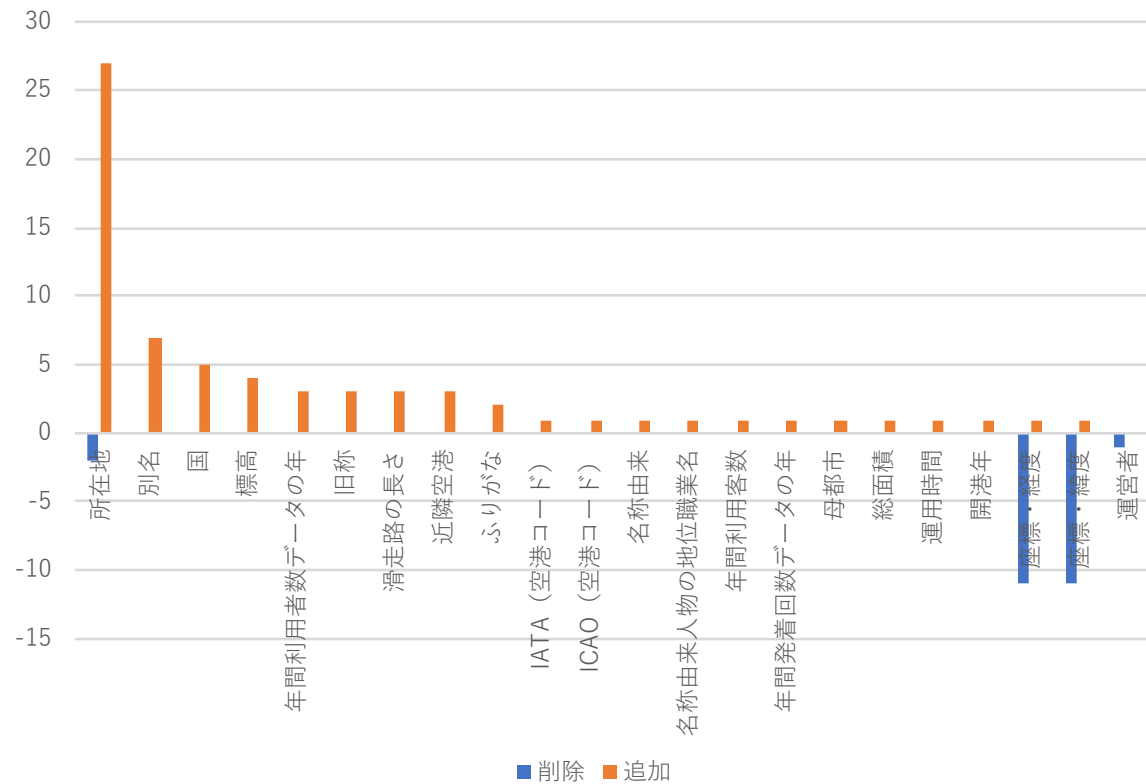


修正された属性値数 (市区町村名)

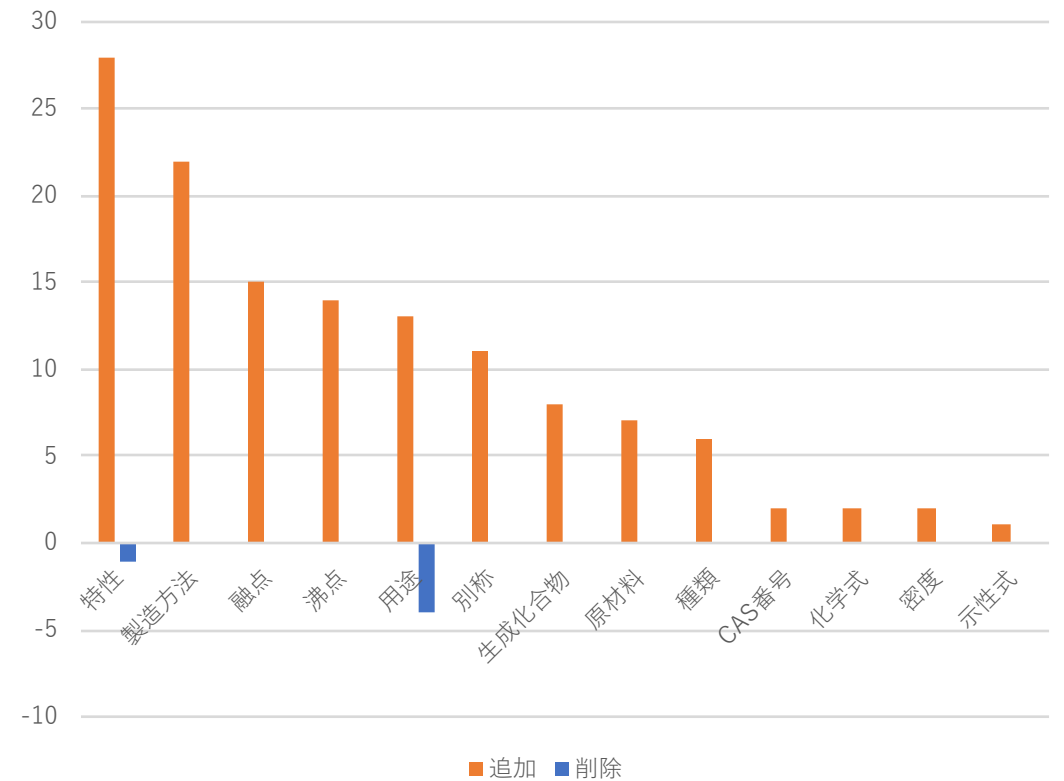


# 正解データの修正結果(2/2)

修正された属性値数（空港名）



修正された属性値数（化合物名）



# まとめ

- 参加チーム：7 (+1) チーム
- 結果の比較
  - **市区町村名・空港名・人名**は豊技大、**企業名・化合物名**はユニシスが最高性能
  - 参考結果も含めると**人名**はユニシス、**市区町村名**はAIPが最高性能
  - 属性値がInfoboxに多く記述され、Infobox種類数が少ないカテゴリ：空港名
    - パターンベースの手法（豊技大）が有効
  - 属性値が記事全体にされるカテゴリ：人名、企業名、市区町村名
    - 機械学習ベースの手法（ユニシス、富士ゼロックス、凸版印刷）が有効
- Infoboxに出現しづらい属性はまだ改善の余地有り
  - 列挙が必要な属性、フレーズの抽出が必要な属性が課題
  - 結果の総和で見ると再現率が高い属性が多い
    - Ensemble LearningによるKB構築コスト低減に期待できる
    - トレーニングデータの増量ができそう
- 空港名は完了（予定）
  - 全体で1600件ほどなので人手による提出結果のチェックが可能