

# Top Level Categories and Attributes for Entity Representation

Barbara Bazzanella, Paolo Bouquet, Heiko Stoermer

2008

# Contents

<b>1</b>	<b>Top-level Categories</b>	<b>8</b>
<b>2</b>	<b>Entity Identification Experiment</b>	<b>10</b>
<b>3</b>	<b>List of the main attributes for each category</b>	<b>13</b>
3.1	Person . . . . .	14
3.2	Organization . . . . .	20
3.3	Event . . . . .	25
3.4	Artifact . . . . .	31
<b>4</b>	<b>Location</b>	<b>37</b>
4.1	Sharedness and Distinctiveness . . . . .	42
4.2	Attributes Relevance . . . . .	46
4.3	A Rank-based Model of Relevance . . . . .	111
4.3.1	Relevance and Rank-based Relevance: a comparison . .	111
4.4	Final considerations . . . . .	129
	<b>Bibliography</b>	<b>130</b>

## List of Tables

1	Normalization examples . . . . .	13
2	Person (it) . . . . .	15
3	Person (eng) . . . . .	16
4	Person (all) . . . . .	17
5	Person: number of attributes for each class of frequency . . . .	18
6	Person (attributes for each subcategory) . . . . .	19
7	Organization (it) . . . . .	20
8	Organization (eng) . . . . .	21
9	Organization (all) . . . . .	21
10	Organization: number of attributes for each class of frequency	22
11	Organization (attributes for each subcategory) . . . . .	24
12	Event (it) . . . . .	25
13	Event (eng) . . . . .	26
14	Event (all) . . . . .	26
15	Event: number of attributes for each class of frequency . . . .	27
16	Event (attributes for each subcategory) . . . . .	28
17	Artifact (it) . . . . .	31
18	Artifact (eng) . . . . .	32
19	Artifact (all) . . . . .	33
20	Artifact: number of attributes for each class of frequency . . .	34
21	Artifact (attributes for each subcategory) . . . . .	36
22	Location (it) . . . . .	37
23	Location (eng) . . . . .	38
24	Location (all) . . . . .	38
25	Location: number of attributes for each class of frequency . .	39
26	Location (attributes for each subcategory) . . . . .	41
27	Distinctiveness (it) . . . . .	44
28	Distinctiveness (eng) . . . . .	46
29	Relevance Politician (it) . . . . .	48
30	Relevance Politician (eng) . . . . .	49
31	Relevance Manager (it) . . . . .	50
32	Relevance Manager (eng) . . . . .	51
33	Relevance Professor (it) . . . . .	52
34	Relevance Professor (eng) . . . . .	53
35	Relevance Sportsperson (it) . . . . .	54
36	Relevance Sportsperson (eng) . . . . .	55
37	Relevance Actor (it) . . . . .	56
38	Relevance Actor (eng) . . . . .	57
39	Relevance Person (it) . . . . .	58

40	Relevance Person (eng)	60
41	Relevance Company (it)	61
42	Relevance Company (eng)	62
43	Relevance Association (it)	63
44	Relevance Association (eng)	64
45	Relevance University (it)	65
46	Relevance University(eng)	66
47	Relevance Government(it)	67
48	Relevance Government (eng)	68
49	Relevance Agency (it)	69
50	Relevance Agency (eng)	70
51	Relevance Organization (it)	71
52	Relevance Organization (eng)	72
53	Relevance Conference (it)	73
54	Relevance Conference (eng)	74
55	Relevance Meeting (it)	75
56	Relevance Meeting (eng)	76
57	Relevance Exhibition (it)	77
58	Relevance Exhibition (eng)	78
59	Relevance Show (it)	79
60	Relevance Show (eng)	80
61	Relevance Incident (it)	81
62	Relevance Incident (eng)	82
63	Relevance Event (it)	83
64	Relevance Event (eng)	84
65	Relevance Sports Event (it)	85
66	Relevance Sports Event (eng)	86
67	Relevance Product (it)	87
68	Relevance Product (eng)	88
69	Relevance Artwork (it)	89
70	Relevance Artwork (eng)	90
71	Relevance Building (it)	91
72	Relevance Building (eng)	92
73	Relevance Book (it)	93
74	Book (eng)	94
75	Relevance Article of clothing (it)	95
76	Relevance Article of clothing (eng)	96
77	Relevance Object(it)	97
78	Relevance Object (eng)	98
79	Relevance Tourist Location(it)	99
80	Relevance Tourist Location (eng)	100

81	Relevance City (it)	101
82	Relevance City (eng)	102
83	Relevance Shop (it)	103
84	Relevance Shop (eng)	104
85	Relevance Hotel (it)	105
86	Relevance Hotel (eng)	106
87	Relevance Restaurant (it)	107
88	Relevance Restaurant (eng)	108
89	Relevance Place (it)	109
90	Relevance Place (eng)	110
91	Relevance measures: Politician (it)	112
92	Relevance measures: Politician (eng)	113
93	Relevance measures: Manager (it)	113
94	Relevance measures: Manager (eng)	113
95	Relevance measures: Professor (it)	113
96	Relevance measures: Professor (eng)	114
97	Relevance measures: Sportperson (it)	114
98	Relevance measures: Sportperson (eng)	114
99	Relevance measures: Actor (it)	114
100	Relevance measures: Actor (eng)	115
101	Relevance measures: Person (it)	115
102	Relevance measures: Person (eng)	115
103	Relevance measures: Company (it)	115
104	Relevance measures: Company (eng)	116
105	Relevance measures: Association (it)	116
106	Relevance measures: Association (eng)	116
107	Relevance measures: University (it)	117
108	Relevance measures: University (eng)	117
109	Relevance measures: Government (it)	117
110	Relevance measures: Government (eng)	117
111	Relevance measures: Agency (it)	118
112	Relevance measures: Agency (eng)	118
113	Relevance measures: Organization (it)	118
114	Relevance measures: Organization (eng)	118
115	Relevance measures: Conference (it)	119
116	Relevance measures: Conference (eng)	119
117	Relevance measures: Meeting (it)	119
118	Relevance measures: Meeting (eng)	119
119	Relevance measures: Exhibition (it)	120
120	Relevance measures: Exhibition (eng)	120
121	Relevance measures: Show (it)	120

122	Relevance measures: Show (eng)	120
123	Relevance measures: Incident (it)	121
124	Relevance measures: Incident (eng)	121
125	Relevance measures: Event (it)	121
126	Relevance measures: Event (eng)	121
127	Relevance measures: Sports Event (it)	122
128	Relevance measures: Sports Event (eng)	122
129	Relevance measures: Product (it)	122
130	Relevance measures: Product (eng)	122
131	Relevance measures: Artwork (it)	123
132	Relevance measures: Artwork (eng)	123
133	Relevance measures: Building (it)	123
134	Relevance measures: Building (eng)	123
135	Relevance measures: Book (it)	124
136	Relevance measures: Book (eng)	124
137	Relevance measures: Article of clothing(it)	124
138	Relevance measures: Article of clothing(eng)	124
139	Relevance measures: Object (it)	125
140	Relevance measures: Object (eng)	125
141	Relevance measures: Tourist Location (it)	125
142	Relevance measures: Tourist Location (eng)	125
143	Relevance measures: City (it)	126
144	Relevance measures: City (eng)	126
145	Relevance measures: Shop (it)	126
146	Relevance measures: Shop (eng)	126
147	Relevance measures: Hotel (it)	127
148	Relevance measures: Hotel (eng)	127
149	Relevance measures: Restaurant (it)	128
150	Relevance measures: Restaurant (eng)	128
151	Relevance measures: Place (it)	128
152	Relevance measures: Place (eng)	128

## List of Figures

1	Person: attributes listed by more than 5% of subjects in the Italian version of the experiment . . . . .	16
2	Person: attributes listed by more than 5% of subjects in the English version of the experiment . . . . .	17
3	Person: attributes listed by more than 5% of subjects for aggregated data . . . . .	18
4	Organization: attributes listed by more than 5% of subjects in the Italian version of the experiment . . . . .	22
5	Organization: attributes listed by more than 5% of subjects in the English version of the experiment . . . . .	23
6	Organization: attributes listed by more than 5% of subjects for aggregated data . . . . .	23
7	Event: attributes listed by more than 5% of subjects in the Italian version of the experiment . . . . .	29
8	Event: attributes listed by more than 5% of subjects in the English version of the experiment . . . . .	29
9	Event: attributes listed by more than 5% of subjects for aggregated data . . . . .	30
10	Artifact: attributes listed by more than 5% of subjects in the Italian version of the experiment . . . . .	33
11	Artifact: attributes listed by more than 5% of subjects in the English version of the experiment . . . . .	35
12	Artifact: attributes listed by more than 5% of subjects for aggregated data . . . . .	35
13	Location: attributes listed by more than 5% of subjects in the Italian version of the experiment . . . . .	39
14	Location: attributes listed by more than 5% of subjects in the English version of the experiment . . . . .	40
15	Location: attributes listed by more than 5% of subjects for aggregated data . . . . .	40

# 1 Top-level Categories

The first step in our entity representation experiment was to select an appropriate collection of top-level categories to define the type of entities under investigation. In order to obtain a reasonably exhaustive, but at the same time limited, set of categories we adopted a “top-down” approach. We analyzed the main top-level ontologies available in literature (Wordnet [1], Dolce [6, 2], Sumo [5], Cyc [3]) to integrate important ontological distinctions from those ontologies. The goal was identifying, by granularity choice, a set of few categories (neither too abstract nor too specific) to use as a test-bed on which to perform the experimental investigation about how people describe entities. At the end of our analysis we identified the following six top-level categories:

1. Person
2. Organization
3. Event
4. Artifact
5. Location
6. Other

We point out that the last category (Other) is a miscellaneous category that contains all entities that are not classifiable in one of the others categories and can be thought (adopting the formalism of the set theory) such as the complement of a set that is represented by the union of the first five categories.

Another aspect that has to be remarked is about the level of abstraction of our categories. Our choice was guided by two constraints. The first is related to the cognitive reliability of the categories (it is well-known that categories are organized into a hierarchy from the most general to the most specific, but the level that is most cognitively basic is in the middle of the hierarchy [7]). The second is more connected to the assumed use of the final system. Directly connected to the latter constraint is the choice about the first category, Person. Although a more general category, such as Being, allowed us a better ontological coverage, including for example animals, it is hard to believe that this type of entities would populate our system.

As evident from the list we limited our analysis to a subclass of entities that we can describe as “physical” entities (things that have a position in



space/time), missing out “abstract” entities (things that do not have spatial nor temporal qualities, and that are not qualities themselves). The distinction between physical and abstract entities is at the base of the SUMO ontology (physical entity vs. abstract entity), the DOLCE ontology (endurant, perdurant particular vs. quality and abstract particular) and the CYC ontology (Intangible thing vs. Individual thing). The notion of abstraction is also present in WordNet, but has a different ontological coverage, not referring to state, psychological feature, action and phenomenon.

Following the distinction proposed by the CYC Ontology, we can distinguish between temporal entities and spatial entities, which justifies two of our top categories: Event and Location. An Event is a thing that occupies a point (or period) in time, whereas a Location is a thing that occupies a space. Both can have spatial and temporal parts, but the ontological nature is determined only by the *essential* parts that are temporal for events and spatial for locations.

Another important ontological assumption that we followed to build our list of top-level categories is related to the behavior of the entity in time. This distinction is connected to the difference between what philosophers usually call “continuants” and “occurrents”, or using the terminology adopted in the Dolce framework between “endurants” and “perdurants”. The main idea is that there are entities (endurants) that are wholly present (all their parts are present) at any time at which they exist and other entities (perdurants) that extend in time and are only partially present for any time at which they exist because some of their temporal parts may be not present. This fundamental distinction allowed us to distinguish between entities that *are* in time like for example Persons or Artifacts and entities that *happen* in time like Events, keeping another distinction that we can find both in the Sumo ontology (object vs process) and in the Dolce ontology (perdurant vs endurant).

A further ontological distinction we made within our basic categories is related to “agentivity”. This property refers to the attribution of intentions, desires and believes and the ability to act on those intentions, desires and believes. On the basis of this assumption we can distinguish physical entities that are agentive such as Person or groups of several agents operating together like Organization and entities that are not-agentive such as Artifact.

Another difference that is taken into account is that between “Individual” entities and “Collection”. This ontological constrain is evident both in Sumo and CYC, and is used to explain the notion of collective entities such as Organization, whose members can be added and subtracted without thereby changing the identity of the collective. A similar distinction results from WordNet that tells apart Entity, defined as something having concrete

existence (living or non-living) and Group that is any number of entities (members) considered as a unit.

After making explicit the representation of the so-called ontological commitments (abstract vs physical, temporal vs spatial, endurant vs perdurant, agentive vs non-agentive, individual vs collective), we can suggest a brief description of each of our top-level categories.

1. *Person*: a physical entity, endowed with temporal parts that can change as a unit (endurant) and able to express desires, intentions and believes (agent).
2. *Organization*: a physical collective entity, whose members are intelligent agents. In terms of behavior in time, an organization changes in time as a whole object so we can define it an endurant. As a collection of agents that operate together, an organization can be considered an agent entity, characterized from desires, intentions and believes.
3. *Event*: a physical individual entity that happens in time, perdurant.
4. *Artifact*: a physical entity intentionally created by an agent (or a group of agents working together) to serve some purpose or perform some function. An artifact is a non-agentive endurant.
5. *Location*: a physical individual entity that has a spatial extent, endurant.
6. *Other*: all the rest

## 2 Entity Identification Experiment

After establishing the top level categories of our study, we conducted an experiment (Entity Identification Experiment) in order to evaluate how people describe entities belonging to such categories.

The goal of this experiment was investigating which attributes are considered more relevant by people to identify type of entities selected as exemplars of the main categories reported above. In fact, since our top level categories were at an high level of abstraction, we decided to introduce a certain number of subcategories for each of them in addition to the simple top level category (named “neutral category”). There are two main reasons for this choice. The first is justifiable in terms of cognitive relevance. Categories more closed to the basic level are more natural and simple to describe. The second is related

to the aim to investigate potential differences inside to the upper level categories in terms of attributes reported, identifying (in addition of attributes common to all different subcategories) also possible specific attributes for specific subcategories. For each top level category we developed 6 (7 for the category “Event”) different scenarios one for each subcategory including the neutral category. The experiment was conducted with a between-subjects design that is one subject was randomly assigned to only one combination of 5 scenarios (one for each top level category). We asked participants to produce a list of all attributes relevant for the specified category (feature-listing task <sup>1</sup>) in order to obtain a unique profile of the entities that populate that category. We report below the list of all categories of the experiment.

PERSON:

- politician
- manager
- professor
- sportsperson
- actor/actress
- person (neutral category)

ORGANIZATION:

- company
- association
- university
- government
- agency
- organization (neutral category)

EVENT:

---

<sup>1</sup>The feature-listing task is a procedure for empirically deriving semantic feature norms, widely used to test theories of semantic representation that use semantic features as their representational currency (for a detailed explanation of the method see [4])

- conference
- meeting
- exhibition
- show
- incident
- event (neutral category)
- sports event

#### ARTIFACT

- product
- artwork
- building
- book
- article of clothing
- object (neutral category)

#### LOCATION

- tourist location
- city
- shop
- hotel
- restaurant
- place (neutral category)

The experiment was conducted in three different versions: English (eng), Italian (it) and Chinese (chi) and was administered on line.

### 3 List of the main attributes for each category

In this section we report the results of the English and the Italian versions of the experiment<sup>2</sup>.

Before starting the quantitative analysis, we normalized the data at three different levels to reduce the noise in the data set and improve the efficacy of the next phase. The first normalization step was performed to remove all typing errors. The second step was finalized to eliminate the noise due to differences in the use of articles, prepositions, number (singular and plural), word order. Finally the last normalization step considered carefully semantic overlaps and synonym expressions. In the table 1 we report some examples of this preliminary phase.

Attributes	Normalized form	Type of Normalization
Surname	Surname	Typing error (1)
Number of telephone	Phone number	Order (2)
Date of birth	Birth date	Order (2)
near cities	neighbouring cities	Semantic overlapping (3)
Zip code	Post code	Synonyms (3)
kind of sport	type of sport	Synonyms (3)

Table 1: Normalization examples

Completed the normalization phase we got on to the quantitative analysis. In the first step of the quantitative analysis we evaluated the frequency of each attribute<sup>3</sup> for each top level category, without distinguishing between subcategories, but keeping initially distinct the data from the two language versions to identify possible differences. In order to make easier the comprehension of the results for not-Italian speakers we reported the attributes of the Italian version with a indicative translation in English, enclosed in parenthesis. Because of the huge amount of attributes mentioned by subjects (the most part of which are unique attributes) we decided to report only attributes listed at least by 5% of participants. This choice is motivated not only by the intent of simplifying the exposure of the results, but also by the final goal of the experiment that is finding attributes considered relevant by the majority of subjects. However, to give evidence about the extension of the conceptual boundaries, we report the total number of different attributes

---

<sup>2</sup>Because of the limited number of data from the Chinese version we decided to keep this data for qualitative considerations.

<sup>3</sup>number of subjects that reported that attribute for the specified category

listed by subjects for each top level category and the number of attributes for each class of frequency. This measures can be useful on the one hand to give an idea about the richness and complexity of the cognitive representation of the categories, on the other hand to put in evidence the disparity between the set of few attributes with high frequencies and the set of many attributes with low frequencies.

In the second phase of the analysis we considered the distinction between subcategories to identify the most frequent attributes within each of these subcategory.

### 3.1 Person

The table 2 shows the absolute ( $F$ ) and relative ( $f$  and  $f\%$ )<sup>4</sup> frequencies of attributes for the Italian version of the experiment. A graphical representation of the percentage frequencies is showed in figure 1.

The same analysis for the English version is reported in table 3 and in figure 2 respectively. The data about the total number of attributes and the distribution of these for each class of frequency is available in the table 5.

As evident from the tables, the attribute most frequently reported in the both versions is “name”, followed by “age”. Only in the English version we find “gender” as one of the three most relevant attributes. We can note that “tax code” is consider a significant identification attribute only in the Italian version, whereas “email” and “country” stand out only in the English version. It is worth noting as physical attributes (such as “height” and “weight”) are present in the both lists and psychological characteristics (“character”) are also considered important to identify person entity (but this emerges only in the Italian version). Except for these differences we can note a substantial matching between the two versions. To a have a global representation of the phenomenon under investigation we computed the same analysis reported above for the aggregated data of the two versions (see table 4 and figure 3). As mentioned above, a further part of the analysis consisted of a comparison between subcategories. In the table 6 we report the frequencies for the different subcategories, considering only attributes whose frequency is at least 20%. We note that two attributes are subcategory-independent: “name” and “age”, that is they are present in any subcategory. Only for the Italian version “surname” is also present in each subcategory, whereas “gender” is well-represented (4 out of 6) in the English version. There are others attributes that are specific (distinctive) for specific subcategories: “politi-

---

<sup>4</sup>The relative frequency  $f$  is the absolute frequency  $F$  divided by  $N$  that is the number of subjects. If the result is multiplied for 100 we obtain the percentage frequency  $f\%$

cal view” and “party” for *politician*, “role” and “experience” for *manager*, “school matter”, “publication”, “educational institution”, “email”, “research area” and “education” for *professor*, “level” and “sport specialty” for *sports person*, “films” for *actor/actress*.

Another consideration is about the neutral category (*person*) that shows many attributes in common with the general results. In the Italian version “tax code” is the distinctive attribute of the neutral category.

<b>Attributes</b>	F	f
nome (name)	89	0.52
età (age)	73	0.42
cognome (surname)	64	0.37
tipo (type)	56	0.32
data di nascita (birth-date)	34	0.19
esperienze (experiences)	29	0.17
titolo di studio (education)	27	0.15
ruolo/i (role/s)	24	0.14
luogo di nascita (birth-place)	21	0.12
nazionalità (nationality)	20	0.12
specialità sportiva (sport specialty)	20	0.11
residenza (residence)	17	0.10
codice fiscale (tax code)	16	0.09
materia di insegnamento (course of teaching)	16	0.09
orientamento politico (political view)	15	0.09
professione (profession)	15	0.09
altezza (height)	13	0.08
partito (party)	13	0.08
carattere (character/nature)	10	0.06
films (films)	10	0.06
peso (weight)	10	0.06
azienda di appartenenza (company)	9	0.05
N=171		

Table 2: Person (it)

Attributes	F	f
name	110	0.75
age	49	0.33
gender	44	0.30
birth-day	29	0.2
party	24	0.16
surname	24	0.16
education	22	0.15
country	20	0.14
nationality	19	0.13
type of sport/s	18	0.12
email	16	0.11
position/s held	16	0.11
address	13	0.10
department	13	0.09
height	13	0.09
phone number	13	0.09
birth-place	12	0.08
role	12	0.08
university	11	0.08
occupation	10	0.08
weight	10	0.07
title/s	10	0.07
experience	9	0.07
films	8	0.05
awards	8	0.05
N=145		

Table 3: Person (eng)

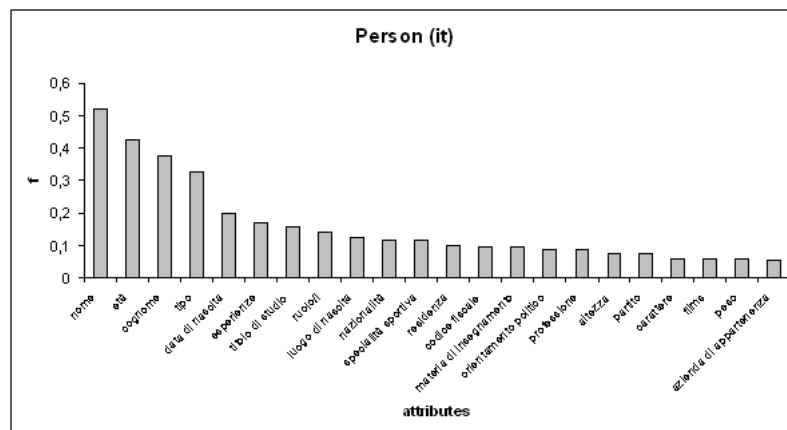


Figure 1: Person: attributes listed by more than 5% of subjects in the Italian version of the experiment



Attributes	F	f
name	199	0.62
age	122	0.38
surname	88	0.27
birth-date	63	0.19
type	57	0.18
education	50	0.15
gender	44	0.15
nationality	39	0.12
experience	38	0.12
party	37	0.11
role/s	35	0.11
type of sports	34	0.10
birth-place	33	0.10
height	26	0.08
occupation	25	0.07
country	24	0.07
email	23	0.07
weight	20	0.06
residence	19	0.06
films	18	0.05
political view	18	0.05
phone number	18	0.05
address	17	0.05
tax code	16	0.05
position/s held	16	0.05
N=316		

Table 4: Person (all)

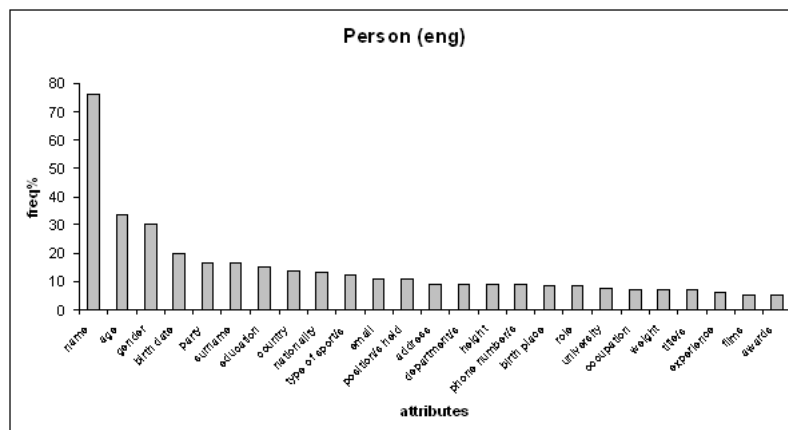


Figure 2: Person: attributes listed by more than 5% of subjects in the English version of the experiment

Italian version		English version	
Frequency	N. Attributes	Frequency	N. Attributes
89	1	110	1
73	1	49	1
64	1	44	1
56	1	29	1
34	1	24	2
29	1	22	1
27	1	20	1
24	1	19	1
21	1	18	1
20	2	16	2
17	1	13	4
16	2	12	2
15	2	11	1
13	2	10	3
10	3	9	1
9	1	8	2
8	2	7	2
7	6	6	1
6	6	5	9
5	4	4	6
4	8	3	13
3	16	2	30
2	33	1	288
1	174		
N. Different Attributes	271	N. Different Attributes:	376

Table 5: Person: number of attributes for each class of frequency

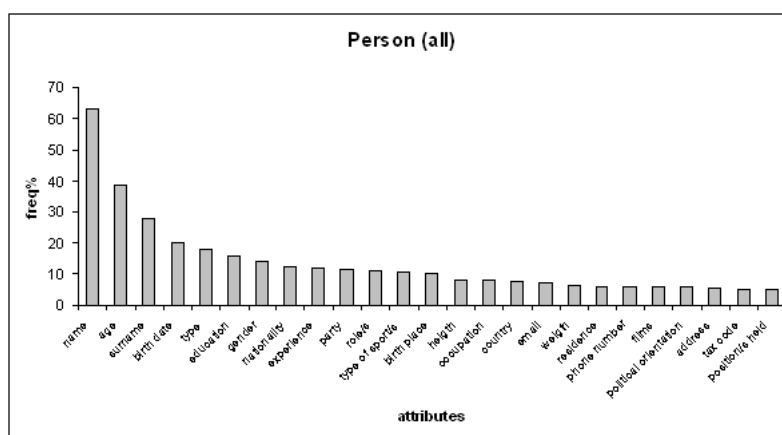


Figure 3: Person: attributes listed by more than 5% of subjects for aggregated data

Subcategory	Attributes (it)	f	Attributes (eng)	f
<i>Politician</i>	età (age)	0.56	party	0.8
	nome (name)	0.46	name	0.63
	orientamento politico (political view)	0.46	age	0.43
	partito (party)	0.43	country	0.33
	cognome (surname)	0.36	gender	0.33
	tipo (type)	0.36	position/s held	0.26
	ruolo (role)	0.33		
	titolo di studio (education)	0.30		
	esperienze (experiences)	0.23		
<i>Manager</i>	nome (name)	0.46	name	0.76
	cognome (surname)	0.39	age	0.33
	azienda (company)	0.28	department	0.23
	età (age)	0.25	experience	0.23
	ruolo (role)	0.21		
	tipo (type)	0.21		
	titolo di studio (education)	0.21		
<i>Professor</i>	nome (name)	0.52	name	0.87
	materia di insegnamento (school matter)	0.64	university	0.41
	età (age)	0.36	department/s	0.33
	cognome (surname)	0.32	education	0.29
	istituzione in cui insegna (educational institution)	0.24	publications	0.29
	pubblicazioni (publications)	0.20	age	0.20
	tipo (type)	0.20	email	0.20
			research area	0.20
			surname	0.20
<i>Sportsperson</i>	specialità sportiva (sport specialty)	0.53	name	0.63
	età (age)	0.46	type of sport	0.6
	nome (name)	0.46	age	0.33
	tipo (type)	0.30	gender	0.3
	cognome (surname)	0.23	birth-date	0.23
	data di nascita (birthdate)	0.20		
	livello (level)	0.20		
<i>Actor/actress</i>	età (age)	0.51	name	0.88
	tipo (type)	0.51	birth date	0.38
	nome (name)	0.48	films	0.38
	esperienze (experiences)	0.45	gender	0.33
	nazionalità (nationality)	0.35	country	0.27
	cognome (surname)	0.32	age	0.22
	films (films)	0.32		
	data di nascita (birth date)	0.22		
	titolo di studio	0.22		
<i>Person</i> neutral category	nome (name)	0.74	name	0.73
	cognome(surname)	0.62	gender	0.53
	data di nascita (birth-date)	0.37	birth date	0.42
	età (age)	0.37	age	0.38
	luogo di nascita (birth-place)	0.37	education	0.23
	codice fiscale (tax code)	0.29	height	0.23
	professione (profession)	0.29	nationality	0.23
	altezza (height)	0.25	occupation	0.23
	residenza (residence)	0.25	surname	0.23
	tipo (type)	0.25		
	carattere (character)	0.22		
	peso (weight)	0.22		

Table 6: Person (attributes for each subcategory)

## 3.2 Organization

Symmetrically to the previous analysis, the table 7, 8 and 9 and figures 4, 5, 6 report the frequency of attributes for the top level category *organization*. Once again “name” is the attribute more frequent regardless of the kind of version. In the top part of the all lists we can find attributes that specify information about localization, such as “location”, “address”, “city” and “country”. Considering the subcategory-based analysis, the only attribute that is common to all subcategories (in both versions) is “name” but a substantial overlapping between attributes of different subcategories is evident. For example “location” and “type” are two attributes shared by almost all subcategories. In particular we point out the strong similarity between attributes of *company* and *agency* subcategories (“name”, “location”, “type”, “address” and “number of employees” are all common attributes). The presence of a consistent number of common attributes comes with a spare number of attributes specific for one or few subcategories. Probably the subcategory that contains more unique attributes is *government* but this seems to be the most difficult to describe as is evident from the low number of attributes listed by subjects.

<b>Attributes</b>	<b>F</b>	<b>f</b>
nome (name)	87	0.51
tipo (type)	54	0.32
luogo (location)	45	0.26
scopo/i (aim/s)	44	0.26
settore (sector)	23	0.13
sede (headquarters)	19	0.11
numero dipendenti (number of employees)	18	0.10
indirizzo (address)	15	0.09
attività (activity)	12	0.07
data di fondazione (foundation date)	14	0.07
dimensione (dimensions)	11	0.06
servizi (services)	11	0.06
città (city)	10	0.06
stato (country)	10	0.06
nazione (nation/country)	10	0.05
fatturato (turnover)	9	0.05
orientamento politico (political view)	9	0.05
partita iva (VAT number)	9	0.05
facoltà (faculty)	9	0.05
N=168		

Table 7: Organization (it)

<b>Attributes</b>	<b>F</b>	<b>f</b>
name	77	0.56
location	37	0.27
country	34	0.24
address	31	0.22
type	23	0.16
objective/s	16	0.11
web site url	14	0.10
city	13	0.9
number of employees	13	0.09
size	12	0.08
date of foundation	11	0.08
business type	10	0.07
members	10	0.07
number of students	8	0.06
phone number/s	8	0.06
state	8	0.06
faculty/ies	8	0.06
head	7	0.05
department/s	7	0.05
role	7	0.05
contact	7	0.05
N=137		

Table 8: Organization (eng)

<b>Attributes</b>	<b>F</b>	<b>f</b>
name	164	0.53
location	81	0.25
type	77	0.25
objective/s	52	0.17
address	46	0.15
country	42	0.13
number of employees	38	0.12
city	23	0.07
date of foundation	23	0.07
size	23	0.07
web site url	21	0.06
state	19	0.06
headquarters	19	0.06
members	18	0.05
activity	17	0.05
faculty/ies	16	0.05
N=305		

Table 9: Organization (all)

Italian version		English version	
Frequency	N. Attributes	Frequency	N. Attributes
87	1	77	1
54	1	37	1
45	1	34	1
43	1	31	1
23	1	23	2
19	1	14	1
18	1	13	2
15	1	12	1
14	1	11	1
12	2	10	2
11	2	8	4
10	3	7	4
9	4	6	3
8	2	5	7
7	4	4	6
6	4	3	6
5	5	2	39
4	8	1	269
3	20		
2	40		
1	227		
N. Different Attributes	330	N. Different Attributes:	351

Table 10: Organization: number of attributes for each class of frequency

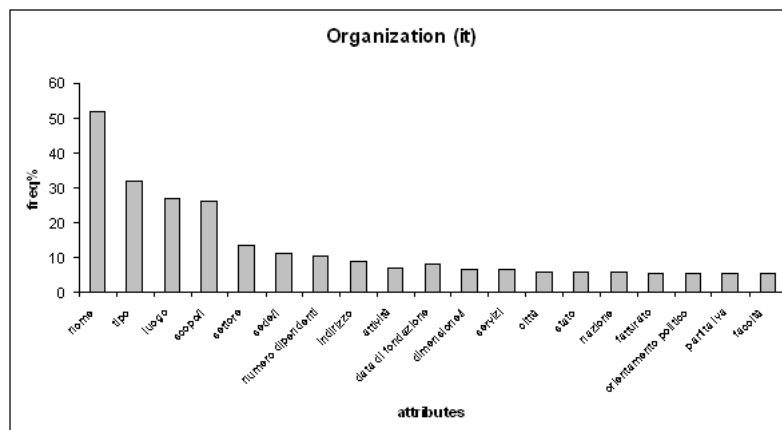


Figure 4: Organization: attributes listed by more than 5% of subjects in the Italian version of the experiment

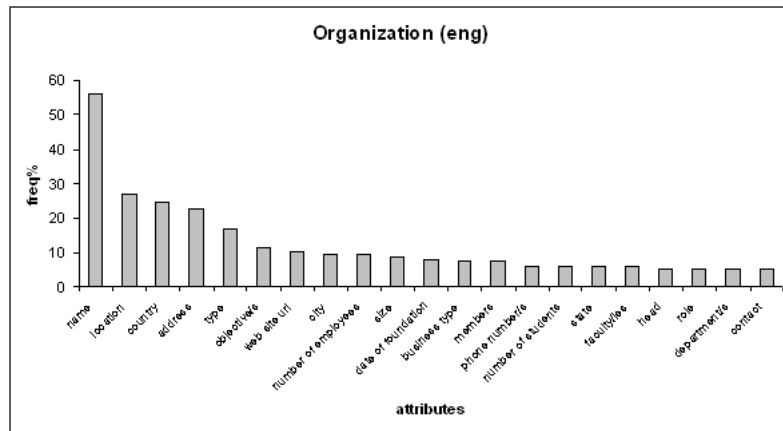


Figure 5: Organization: attributes listed by more than 5% of subjects in the English version of the experiment

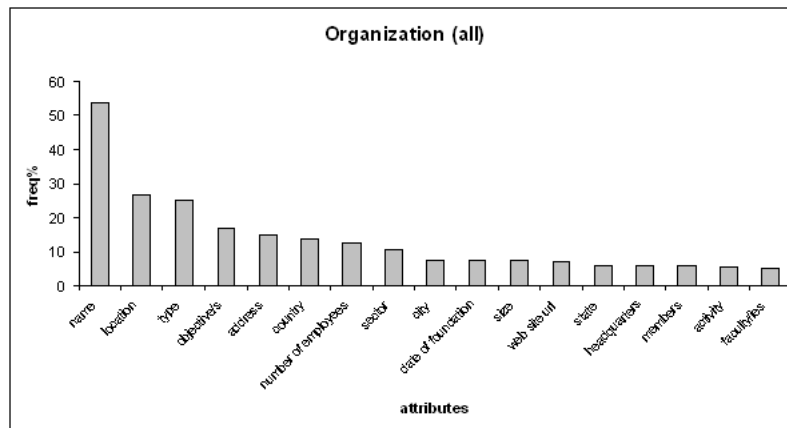


Figure 6: Organization: attributes listed by more than 5% of subjects for aggregated data

Subcategory	Attributes (it)	f	Attributes (eng)	f
<i>Agency</i>	nome (name)	0.68	name	0.83
	luogo (location)	0.34	address	0.38
	tipo (type)	0.31	location	0.38
	numero di dipendenti (number of employees)	0.28	country	0.33
			business type	0.22
			number of employees	0.22
			web site url	0.22
<i>Association</i>	nome (name)	0.60	name	0.52
	scopo/i (objective/s)	0.57	location	0.32
	tipo (type)	0.39	objective/s	0.24
	associati (members)	0.28	type	0.24
	settore (sector)	0.21	website url	0.24
			activity	0.20
			address	0.20
			date of foundation	0.20
			memebers	0.20
<i>University</i>	luogo (location)	0.48	name	0.61
	nome (name)	0.45	location	0.46
	facoltà(type)	0.25	address	0.30
	corsi (sector)	0.22	city	0.26
			number of students	0.26
		country	0.23	
<i>Government</i>	orientamento politico (political orientation)	0.34	country	0.55
	tipo (type)	0.26	name	0.25
	nazione (nation)	0.21		
<i>Agency</i>	nome (name)	0.66	name	0.65
	tipo (type)	0.54	address	0.35
	luogo (location)	0.29	number of employees	0.25
	indirizzo (address)	0.25	type	0.25
	scopo/i (objective/s)	0.25		
	numero di dipendenti (number of employees)	0.20		
<i>Organization</i> neutral category	nome (name)	0.56	name	0.61
	scopo/i (objective/s)	0.56	location	0.28
	tipo(type)	0.33	type	0.28
	settore (sector)	0.26		
	luogo (location)	0.23		
	sede (head office)	0.20		

Table 11: Organization (attributes for each subcategory)



### 3.3 Event

For the top level category *event* we found that “location” is the attribute most frequently listed by the participants, followed by “date” and “time” both in English and in the Italian version. This result represents a substantial difference compared with the previous two categories (*person* and *organization*) where “name” is the most frequent attribute. In fact, if we consider the aggregated data, “name” is the fourth attribute according to the frequency. “Location” is also the unique common attribute in every lists when we split the data on the basis of the subcategory-type. Few attributes turn out to be specific of particular subcategories, more precisely “people and entity involved” for *incident* and “type of sport” for *sport event*.

Attributes	F	f
luogo (location)	126	0.78
data (date)	74	0.45
tipo (type)	68	0.42
ora (time)	57	0.35
partecipanti (participants)	39	0.24
argomento (topic)	38	0.23
durata (duration)	33	0.20
nome (name)	23	0.14
titolo (title)	22	0.13
ente organizzatore (organizers)	17	0.10
prezzo/i (price/s)	16	0.09
persone coinvolte (people involved)	13	0.08
attori (actors)	11	0.06
obiettivo/i (objective/s)	10	0.06
specialità sportiva (sport specialty)	10	0.06
descrizione (description)	9	0.05
N=161		

Table 12: Event (it)

<b>Attributes</b>	<b>F</b>	<b>f</b>
location	116	0.79
date	69	0.47
time	64	0.43
name	49	0.33
participants	40	0.27
type	26	0.17
duration	18	0.12
topic/s	16	0.10
organizer/s	16	0.10
subject/s	15	0.10
end date	13	0.08
purpose/s	13	0.08
title/s	12	0.08
type of sport	11	0.07
city	8	0.05
country	8	0.05
description	8	0.05
sponsor/s	8	0.05
start time	8	0.05
year	8	0.05
N=146		

Table 13: Event (eng)

<b>Attributes</b>	<b>F</b>	<b>f</b>
location	242	0.78
date	143	0.46
time	121	0.39
type	94	0.30
participants	89	0.28
name	72	0.23
topic/s	54	0.17
duration	51	0.16
title	34	0.11
organizers	33	0.10
objective/s	23	0.07
type of sport	21	0.06
people involved	18	0.05
price/s	17	0.05
description	17	0.05
actors	16	0.05
N=307		

Table 14: Event (all)

Italian version		English version	
Frequency	N. Attributes	Frequency	N. Attributes
126	1	116	1
74	1	69	1
68	1	64	1
47	1	49	1
39	1	40	2
38	1	26	1
33	1	18	1
23	1	16	2
22	1	15	1
17	1	12	1
16	1	11	1
13	1	8	6
11	1	7	1
10	3	6	2
9	1	5	4
8	2	4	5
7	2	3	14
6	3	2	31
5	5	1	216
4	8		
3	12		
2	30		
1	142		
N. Different Attributes	221	N. Different Attributes:	292

Table 15: Event: number of attributes for each class of frequency

Subcategory	Attributes (it)	f	Attributes (eng)	f
<i>Conference</i>	luogo (location)	0.9	location	0.84
	argomento/i (topic/s)	0.7	name	0.57
	partecipanti (participants)	0.5	date	0.34
	data (date)	0.45	organizer/s	0.26
	durata (duration)	0.30		
	relatore/i (speaker/s)	0.30		
	titolo (title)	0.25		
	ente organizzatore (organizers)	0.20		
	obiettivo/i (objective/s)	0.20		
ora (time)	0.20			
<i>Meeting</i>	luogo (location)	0.84	location	0.88
	ora (time)	0.63	time	0.80
	argomento/i (topic/s)	0.57	date	0.64
	partecipanti (participants)	0.52	participants	0.52
	tipo(type)	0.21	topic/s	0.35
	type	0.24		
	name	0.20		
<i>Exhibition</i>	luogo (location)	0.72	location	0.75
	argomento/i (topic/s)	0.54	name	0.5
	data (date)	0.31	time	0.37
	durata (duration)	0.31	date	0.31
	tipi (type)	0.27		
	tipolo (title)	0.27		
<i>Show</i>	tipo (type)	0.75	location	0.52
	luogo (location)	0.72	name	0.52
	data (date)	0.43	age	0.33
	durata (duration)	0.29	actors	0.29
	ora (time)	0.29	time	0.29
	prezzo/i (price/s)	0.27	type	0.23
	titolo (title)	0.24		
	attori (actors)	0.21		
<i>Incident</i>	luogo (location)	0.85	time	0.66
	persone coinvolte (people involved)	0.52	location	0.57
	ora (time)	0.38	date	0.42
	tipo (type)	0.38	participants	0.23
	data (date)	0.33		
	entità coinvolte (entities involved)	0.28		
<i>Event</i> neutral category	data (date)	0.71	location	0.91
	luogo (location)	0.71	time	0.52
	tipo (type)	0.66	date	0.43
	partecipanti (participants)	0.38	participants	0.34
	nome (name)	0.33	name	0.30
	durata (duration)	0.23	type	0.26
	ente organizzatore (organizers)	0.23	purpose/s	0.21
	ora (time)	0.23		
<i>Sport event</i> neutral category	luogo (location)	0.76	location	1
	data (date)	0.52	date	0.58
	specialità sportiva (sport specialty)	0.47	type of sport	0.58
	nome (name)	0.38	time	0.29
	tipo (type)	0.28	duration	0.23
		0.29	name	0.23
		0.29	participants	0.23

Table 16: Event (attributes for each subcategory)

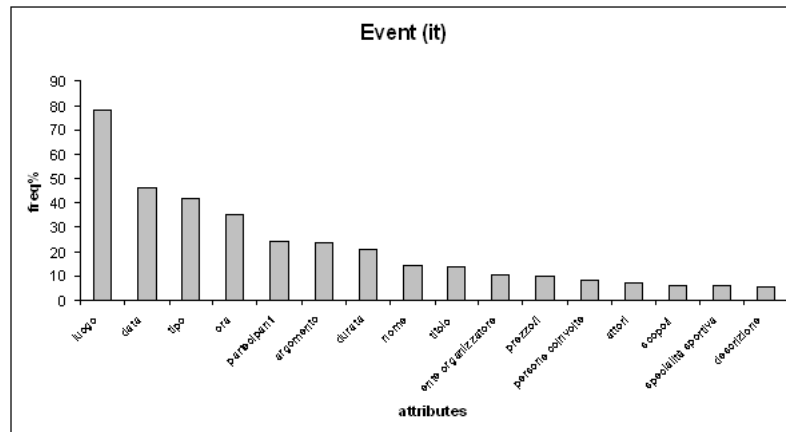


Figure 7: Event: attributes listed by more than 5% of subjects in the Italian version of the experiment

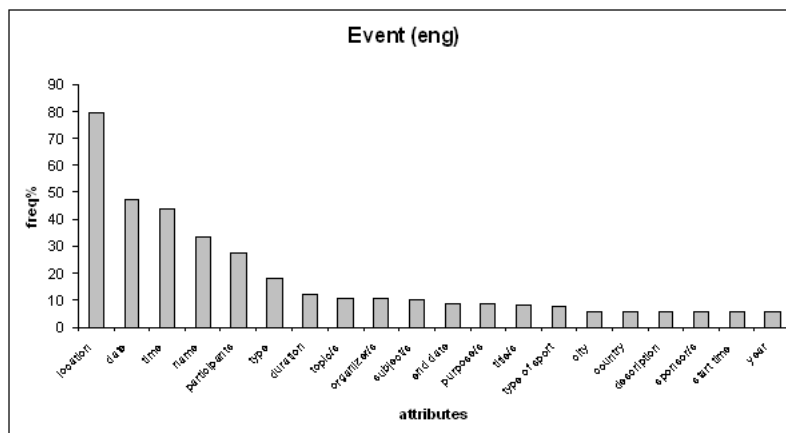


Figure 8: Event: attributes listed by more than 5% of subjects in the English version of the experiment

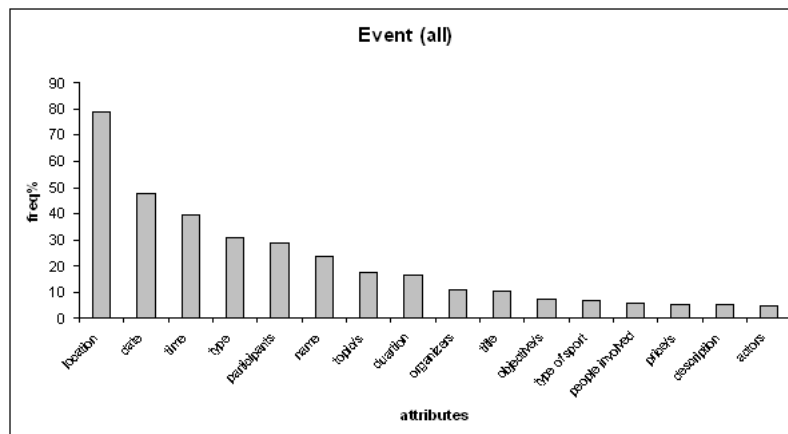


Figure 9: Event: attributes listed by more than 5% of subjects for aggregated data

### 3.4 Artifact

The top level category *artifact* presents an unexpected first attribute: “color” and this result is constant across the two versions of the experiment. Moreover other “physical” attributes seem to be dominant in the descriptions produced by subjects, for example “size”, “material”, “shape”, “height”, “dimensions” and “weight”. It is interesting to note as also for this category “name” is reported after others attributes such as “author” and “type”. In particular “author” is the most frequent attribute for two subcategories: *art-work* and *book*. “Location” appears one of the most important attribute to identify buildings, whereas “price” is the first attribute for *product*. Other attributes are unique for specific subcategories. For example, ISBN is specific for *book* as well as “publisher” and “year of publication”. “Brand” and “size” are considered more relevant to identify “article of clothing” and “shape” appears among the main attributes of “object” that is the neutral category.

Attributes	F	f
colore/i (color/s)	74	0.44
tipo (type)	60	0.35
autore/i (author/s)	51	0.30
dimensione/i (dimension/s)	36	0.21
materiale (material)	35	0.20
prezzo/i (price/s)	34	0.20
titolo (title)	31	0.18
luogo (location)	30	0.17
nome (name)	28	0.16
utilizzo (use)	26	0.15
forma (shape)	22	0.12
taglia/e (size/s)	21	0.12
marca (brand)	20	0.11
numero pagine (number of pages)	16	0.09
peso (weight)	16	0.09
stile (style)	14	0.08
altezza (height)	13	0.07
funzione (function)	13	0.07
editore (editor)	11	0.06
numero di piani (number of floors)	11	0.06
anno di pubblicazione (publication year)	9	0.05
data (date)	9	0.05
N=168		

Table 17: Artifact (it)

<b>Attributes</b>	<b>F</b>	<b>f</b>
color/s	46	0.32
size	35	0.25
name	33	0.23
title/s	29	0.20
type	28	0.20
price/s	25	0.17
author/s	24	0.17
material	24	0.17
location	18	0.12
address	16	0.11
shape	16	0.11
style	16	0.11
use	16	0.11
creation date	14	0.10
ISBN	14	0.10
manufacturer	14	0.10
height	13	0.09
publisher	13	0.09
description	10	0.07
artist/s	9	0.06
owner/s	9	0.06
purpose/s	9	0.06
brand name	8	0.05
category	8	0.05
brand name	8	0.05
category	8	0.05
year	8	0.05
year of publication	8	0.05
weight	7	0.05
number of page	7	0.05
gender intended for	7	0.05
id	7	0.07
N=140		

Table 18: Artifact (eng)



Attributes	F	f
color/s	120	0.38
type	88	0.28
author/s	75	0.24
name	61	0.19
title	60	0.19
material	59	0.19
price/s	59	0.19
size	56	0.18
location	48	0.15
dimension	42	0.13
use	42	0.13
shape	38	0.12
style	30	0.09
brand	28	0.09
height	26	0.08
publisher	24	0.07
address	23	0.07
weight	23	0.07
ISBN	23	0.07
number of pages	23	0.06
creation date	19	0.06
description	18	0.05
year of publication	17	0.05
N=308		

Table 19: Artifact (all)

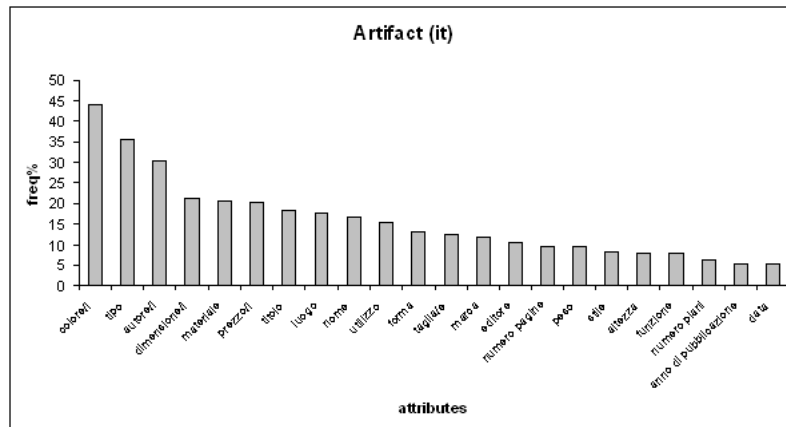


Figure 10: Artifact: attributes listed by more than 5% of subjects in the Italian version of the experiment

Italian version		English version	
Frequency	N. Attributes	Frequency	N. Attributes
74	1	46	1
60	1	35	1
51	1	33	1
36	1	29	1
35	1	28	1
34	1	25	1
31	1	24	2
30	1	18	1
28	1	16	4
26	1	14	3
21	2	13	2
20	1	10	1
16	2	9	3
14	1	8	4
13	2	7	4
11	1	6	4
9	3	5	4
8	3	4	6
7	6	3	14
6	2	2	21
5	7	1	192
4	8		
3	9		
2	33		
1	141		
N. Different Attributes	231	N. Different Attributes:	271

Table 20: Artifact: number of attributes for each class of frequency

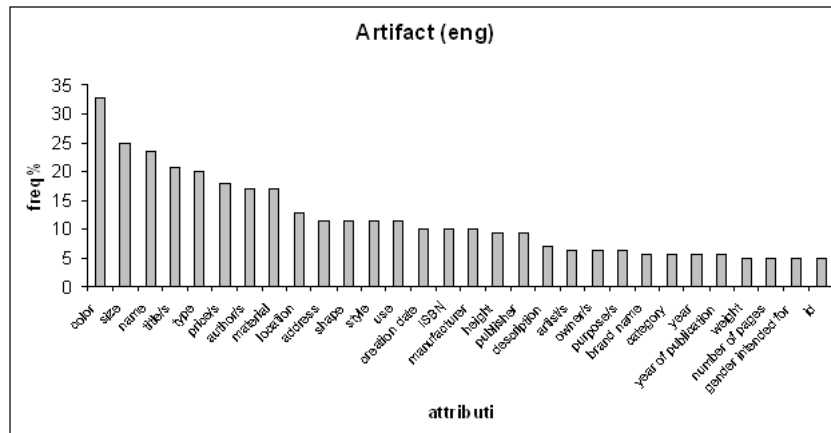


Figure 11: Artifact: attributes listed by more than 5% of subjects in the English version of the experiment

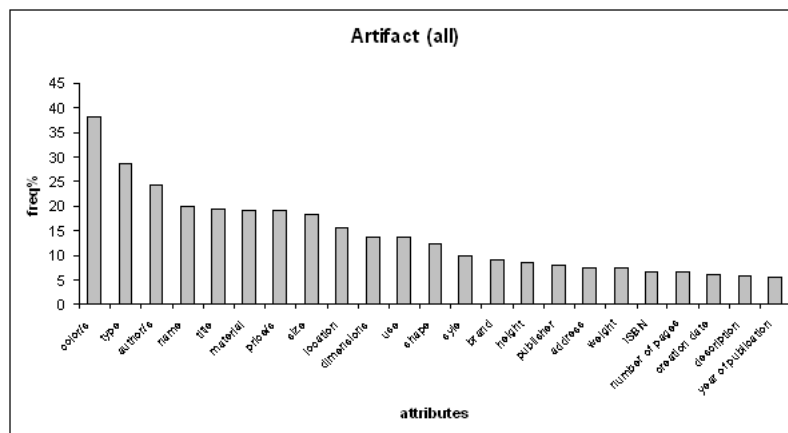


Figure 12: Artifact: attributes listed by more than 5% of subjects for aggregated data

Subcategory	Attributes (it)	f	Attributes (eng)	f
<i>Product</i>	prezzo/i (price/s)	0.6	price/s	0.52
	nome (name)	0.44	name	0.39
	utilizzo (use)	0.44	color	0.21
	tipo (type)	0.40	description	0.21
	colore/i (color/s)	0.36		
	dimensione/i (dimension/s)	0.28		
	caratteristiche (features)	0.20		
	peso (weight)	0.20		
<i>Artwork</i>	autore/i (author/s)	0.76	artist/s	0.47
	luogo (location)	0.4	creation date	0.47
	tipo (type)	0.33	title/s	0.42
	materiale (material)	0.30	material	0.31
	titolo (title)	0.30	style	0.31
	data (date)	0.26	type	0.26
	stile (style)	0.26	date	0.21
	colore/i (color/s)	0.20		
	soggetto (subject)	0.20		
<i>Building</i>	luogo (location)	0.53	address	0.65
	altezza (height)	0.36	location	0.56
	numero di piani (number of floors)	0.36	height	0.47
	colore/i (color/s)	0.33	name	0.34
	dimensione/i (dimension/s)	0.33	color	0.21
	tipo (type)	0.33	architect	0.26
	indirizzo (address)	0.23	number of floors	0.21
	destinatari (recipients)	0.20	owner/s	0.21
	metratura (area mq)	0.20	type	0.21
	utilizzo (use)	0.20		
<i>Book</i>	autore/i (author/s)	0.9	author/s	0.8
	titolo (title)	0.73	title	0.76
	editore (publisher)	0.6	ISBN	0.52
	numero di pagine (number of page)	0.53	publisher	0.52
	anno di pubblicazione (year of publication)	0.30	year of publication	0.32
	tipo (type)	0.20	year	0.20
	ISBN	0.23		
	argomento (topic)			
<i>Article of clothing</i>	colore/i (color/s)	0.87	color	0.72
	taglia/e (size/s)	0.67	size	0.52
	tipo (type)	0.61	type	0.44
	marca (brand)	0.48	material	0.36
	prezzo/i (price/s)	0.38	gender	0.28
	tessuto/i (fabric)	0.25	price/s	0.28
			style	0.28
		brand name	0.20	
<i>Object</i>	colore/i (color/s)	0.90	color	0.56
	materiale (material)	0.63	size	0.48
	forma (shape)	0.63	shape	0.40
	dimensione/i (dimension/s)	0.54	name	0.24
	peso (weight)	0.45	purpose/s	0.20
	funzione (function)	0.40		
	nome (name)	0.40		
	utilizzo (use)	0.27		
	epoca (period)	0.09		

Table 21: Artifact (attributes for each subcategory)

## 4 Location

As seen for *person* and *organization*, also for *location* “name” is the attribute with the highest frequency when we consider the aggregated data. In the Italian version this attribute is preceded by “location”. A part of other attributes which specify information about geographical localization such as “country”, “address”, “region” etc., constant is the presence of “geographical coordinates” expressed as latitude and longitude that is the first attribute for the neutral category. “Number of citizens” seems to be specific for *city* and “amenities” for *tourist location*. “Number of stars” is unique for *hotel* as well as “number of rooms” and “services”. Finally we find “type of cuisine” only in the subcategory *restaurant*.

Attributes	F	f
luogo (location)	78	0.46
nome (name)	73	0.43
tipo (type)	57	0.33
coordinate geografiche (latitudine, longitudine) (geo coordinates, latitude, longitude)	35	0.20
numero abitanti (número of citizens)	29	0.17
indirizzo (address)	28	0.16
prezzo/i (price/s)	19	0.11
servizi (services)	19	0.11
stato (state)	19	0.11
orario (time)	18	0.10
regione (region)	17	0.10
dimensione/i (dimension/s)	16	0.09
merce trattata (products)	14	0.08
categoria di appartenenza (category)	13	0.07
città (city)	13	0.07
numero di stelle (number of stars)	13	0.07
nazione (country)	12	0.07
posizione geografica (geographical position)	11	0.06
clima (climate)	10	0.05
numero di stanze (number of rooms)	10	0.05
attrazioni (amenities)	9	0.05
numero di dipendenti (number of employees)	9	0.05
N=169		

Table 22: Location (it)

<b>Attributes</b>	<b>F</b>	<b>f</b>
name	86	0.59
country	50	0.34
location	48	0.33
address	47	0.32
geo coordinates (latitude, longitude)	43	0.29
city	31	0.21
price/s	22	0.15
type of cuisine	20	0.10
state	15	0.11
population	14	0.09
type	11	0.07
street	10	0.06
post code	10	0.06
owner/s	9	0.06
gps coordinates	8	0.05
size	8	0.05
N=145		

Table 23: Location (eng)

<b>Attributes</b>	<b>F</b>	<b>f</b>
name	159	0.50
location	126	0.40
geo coordinates (latitude, longitude)	78	0.24
address	78	0.23
type	69	0.21
price/s	62	0.19
country	62	0.19
population	43	0.13
city	42	0.13
state	34	0.10
services	26	0.08
dimensions	21	0.08
region	21	0.06
type of cuisine	20	0.06
time	20	0.06
number of stars	16	0.05
N=314		

Table 24: Location (all)

Italian version		English version	
Frequency	N. Attributes	Frequency	N. Attributes
78	1	86	1
73	1	50	1
57	1	48	1
35	1	47	1
29	1	43	1
28	1	31	1
19	3	22	1
18	1	20	1
17	1	15	1
16	1	14	1
14	1	12	1
13	3	10	2
12	1	9	1
11	1	8	2
10	2	7	5
9	2	6	2
7	3	5	4
6	3	4	8
5	7	3	15
4	6	2	26
3	16	1	235
2	38		
1	161		
N. Different Attributes 256		N. Different Attributes: 311	

Table 25: Location: number of attributes for each class of frequency

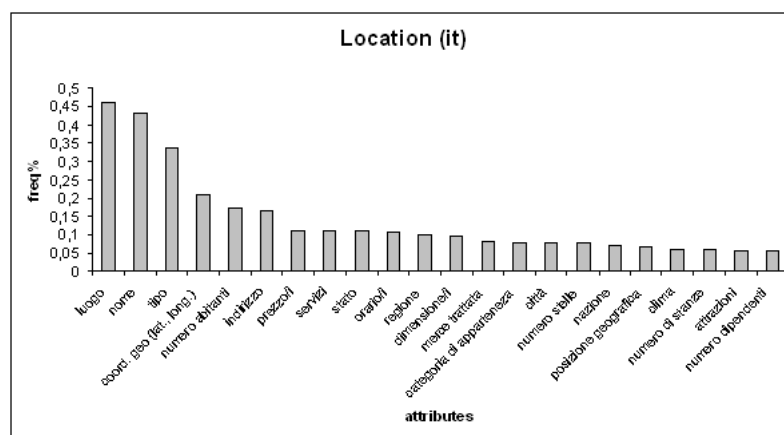


Figure 13: Location: attributes listed by more than 5% of subjects in the Italian version of the experiment

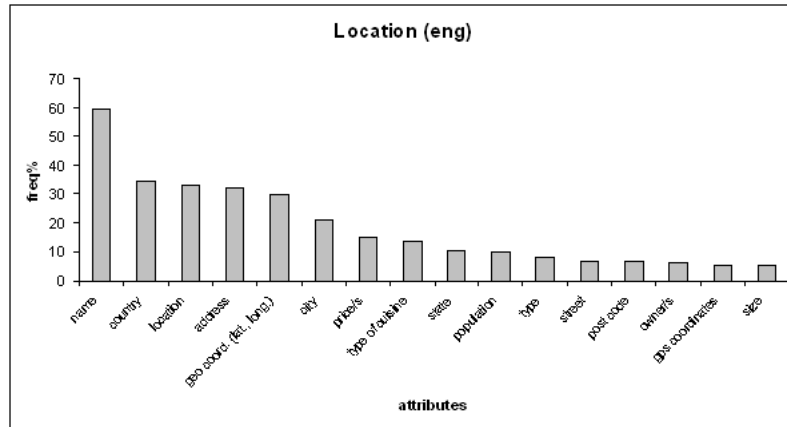


Figure 14: Location: attributes listed by more than 5% of subjects in the English version of the experiment

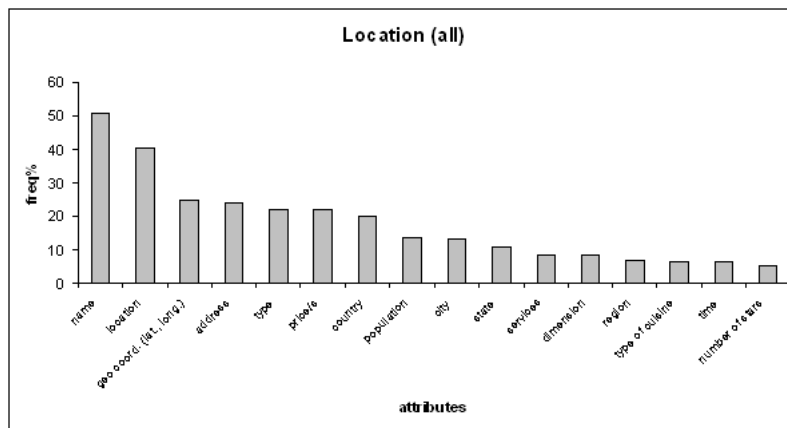


Figure 15: Location: attributes listed by more than 5% of subjects for aggregated data



Subcategory	Attributes (it)	f	Attributes (eng)	f	
<i>Tourist Location</i>	luogo (place)	0.58	name	0.57	
	attrazioni (amenities)	0.33	country	0.42	
	nome (name)	0.29	geo coordinates	0.36	
	tipo (type)	0.29	attractions	0.21	
	numero di abitanti (population)		0.16	city	0.21
				location	0.21
			price/s	0.21	
<i>City</i>	numero di abitanti (number of citizens)	0.8	country	0.52	
	coordinate geografiche (geo coordinates)	0.43	name	0.52	
	nome (name)	0.4	population	0.44	
	stao (state)	0.36	location	0.40	
	regione (region)	0.33	geo coordinates	0.28	
	nazione (country)	0.23	language/s	0.20	
	clima (climate)	0.20			
<i>Shop</i>	luogo (location)	0.55	location	0.42	
	nome (name)	0.52	name	0.38	
	tipo (type)	0.33	height	0.47	
	merce trattata (products sold)	0.33	address	0.33	
	indirizzo (address)	0.33	type	0.23	
	orario/i (timetable)	0.30			
<i>Hotel</i>	luogo (location)	0.66	name	0.83	
	nome (name)	0.55	address	0.5	
	servizi (services)	0.40	location	0.37	
	numero di stanze (number of rooms)	0.37	country	0.29	
	numero di stelle (number of stars)	0.37	city	0.25	
	categoria di appartenenza (category)	0.29			
	indirizzo ( address)	0.22			
<i>Restaurant</i>	tipo (type)	0.7	name	0.75	
	luogo (location)	0.56	address	0.71	
	nome (name)	0.43	type of cuisine	0.67	
	prezzo/i (price/s)	0.33	location	0.46	
	indirizzo (address)	0.30	price/s	0.39	
	orario/i (timetable)		0.23	city	0.28
				country	0.21
<i>Place</i>	coordinate geografiche (geo coordinates)	0.72	geo coordinates	0.92	
	nome (name)	0.31	country	0.5	
	luogo (location)		0.27	name	0.46
			city	0.35	

Table 26: Location (attributes for each subcategory)

## 4.1 Sharedness and Distinctiveness

The analysis of the frequencies reported above evaluates the importance of an attribute  $f$  for a specific category (or subcategory)  $c$ . In this sense it represents a local measure (that we can name “dominance”) of each attribute for a specific category. We can formalize the function of Dominance  $\phi: C \times F \rightarrow N$  in the following way:

$$\text{dominance} = \phi(c, f) = |\{s \in S : f \in F_s^c\}|$$

where  $S$  is the sample of subjects and  $F_s^c$  is the set of attributes listed by the subject  $s$  given the category  $c$ . In other words, the dominance  $\phi$  of the attribute  $f$  for the category  $c$  is the cardinality<sup>5</sup> of the set constituted by all subjects that reported the attribute  $f$  for the category  $c$ . The dominance presents high scores when the attribute is frequently mentioned by subjects in defining the category.

A second dimension of an attribute that we can analyze is the “distinctiveness” of that attribute relative to all (sub)categories. When an attribute is used only in identifying one or few categories its distinctiveness is high, whereas when it is used for many categories (or all) the distinctiveness score is low. The distinctiveness can be calculated as a function  $\psi_d(f) : F \rightarrow [0, 1]$

$$\text{distinctiveness} = \psi_d(f) = 1 - \psi_s(f)$$

where  $\psi_s(f)$  is a function of sharedness  $\psi_s(f) : F \rightarrow [0, 1]$

$$\text{sharedness} = \psi_s(f) = \frac{|C[f]|}{|C|}$$

where  $|C[f]|$  is the collection of the categories that have in common the attribute  $f$  and  $|C|$  is the collection of all categories. If an attribute  $f$  is listed for all categories  $\psi_d(f)$  is 0 and  $\psi_s(f)$  is 1.

The distinctiveness is a global measure because is transversal to all categories and in this sense it is category-independent and frequency-independent. This means that if we consider two different attributes  $f_1$  and  $f_2$  one used by all subject only in the category  $c_1$  and the other used by only one subject only in the category  $c_2$ , their distinctiveness is identical ( $\psi_d(f_1) = \psi_d(f_2) = 1/|C|$ ) regardless of the category and the number of subjects.

The tables 27 and 28 show the sharedness  $\psi_s(f)$  and distinctiveness  $\psi_d(f)$  scores of the attributes above the threshold of 5% in terms of frequency. We point out that to compute these measures we considered the attribute “name” distinct for the five top level categories. This choice, partially suggested by

---

<sup>5</sup>number of the elements of a set

the answers of subjects <sup>6</sup>, seems recommendable in the light of the semantic nuance that the attribute “name” can assume in different contexts. It’s difficult believe that “name” could have the same semantic value for a person than for an organization or a product <sup>7</sup> and this is the same for the other categories.

Attributes	Sharedness $\psi_s(f)$	Distinctiveness $\psi_d(f)$
nome della persona	0.19	0.81
nome dell’organizzazione	0.19	0.81
nome dell’evento	0.16	0.84
nome dell’artefatto	0.19	0.81
nome del luogo	0.19	0.81
età	0,29	0.71
cognome	0,19	0.81
tipo	1	0
data di nascita	0.19	0.81
esperienze	0.16	0.84
titolo di studio	0.19	0.81
ruolo/i	0.22	0.78
luogo di nascita	0.16	0.84
nazionalità	0.29	0.71
residenza	0.19	0.81
specialità sportiva	0.06	0.94
codice fiscale	0.25	0.75
orientamento politico	0.13	0.87
professione	0.10	0.90
altezza	0.23	0.77
partito	0.06	0.94
materia di insegnamento	0.03	0.97
carattere	0.13	0.87
films	0.03	0.97
peso	0.16	0.84
azienda di appartenenza	0.06	0.94
scopo/i	0.48	0.52
luogo	0.77	0.23
settore	0.32	0.68
sede/i	0.26	0.74
numero dipendenti	0.19	0.81
indirizzo	0.55	0.45
attività	0.45	0.55
data di fondazione	0.19	0.81
dimensione/i	0.52	0.48

<sup>6</sup>for example in the Italian version a substantial number of subjects used the attribute “ragione sociale (business name)” instead of “name”

<sup>7</sup>for example there are legal restrictions that prevent the use of the same name for different products. In this sense “name” is an attribute more unambiguous for products than for people.

città	0.55	0.45
stato	0.55	0.45
fatturato	0.13	0.87
partita iva	0.23	0.77
data	0.29	0.71
ora	0.23	0.77
partecipanti	0.19	0.81
argomento	0.16	0.84
durata	0.19	0.81
titolo	0.19	0.81
ente organizzatore	0.29	0.71
prezzo/i	0.39	0.61
persone coinvolte	0.10	0.90
attori	0.10	0.90
specialità sportiva	0.06	0.94
descrizione	0.42	0.58
colore/i	0.29	0.71
autore/i	0.19	0.81
materiale	0.19	0.81
utilizzo	0.13	0.87
forma	0.19	0.81
taglia/e	0.03	0.97
marca	0.10	0.90
editore	0.03	0.97
numero pagine	0.03	0.97
stile	0.10	0.90
funzione	0.10	0.90
anno di pubblicazione	0.03	0.97
numero piani	0.06	0.94
coord. geo (lat., long.)	0.16	0.84
numero abitanti	0.06	0.96
servizi	0.26	0.74
regione	0.19	0.81
merce trattata	0.04	0.96
categoria di appartenenza	0.20	0.80
numero stelle	0.06	0.94
posizione geografica	0.13	0.87
clima	0.10	0.90
numero di stanze	0.04	0.96
attrazioni	0.06	0.94

Table 27: Distinctiveness (it)

Attributes	Sharedness $\psi_s(f)$	Distinctiveness $\psi_d(f)$
name of person	0.19	0.81
name of organization	0.19	0.81

name of event	0.23	0.77
name of artifact	0.19	0.81
name of location	0.19	0.81
age	0.42	0.58
gender	0.19	0.81
birth date	0.19	0.81
party	0.06	0.94
surname	0.19	0.81
education	0.26	0.74
country	0.84	0.16
nationality	0.26	0.74
type of sport/s	0.06	0.94
email	0.39	0.61
position/s held	0.16	0.84
address	0.74	0.26
department/s	0.16	0.84
height	0.26	0.54
phone number/s	0.48	0.52
birth place	0.16	0.84
role	0.35	0.65
university/ies	0.03	0.97
occupation	0.10	0.90
weight	0.29	0.71
title/s	0.48	0.52
experience	0.10	0.90
films	0.06	0.94
location	0.81	0.29
type	0.71	0.32
web site url	0.55	0.45
city	0.58	0.42
number of employees	0.23	0.77
size	0.58	0.42
date of foundation	0.16	0.84
business type	0.13	0.87
members	0.10	0.90
number of students	0.06	0.94
objective/s	0.19	0.81
state	0.39	0.61
head	0.10	0.90
date	0.29	0.71
time	0.29	0.71
participants	0.29	0.71
duration	0.23	0.77
topic/s	0.13	0.87
organizer/s	0.16	0.84
end date	0.19	0.84
purpose/s	0.26	0.90
type of sport	0.06	0.94
description	0.35	0.65

sponsor/s	0.13	0.87
start time	0.19	0.81
subject	0.19	0.81
year	0.26	0.74
color	0.39	0.61
price/s	0.29	0.71
author/s	0.10	0.90
material	0.16	0.84
shape	0.19	0.81
style	0.10	0.90
creation date	0.13	0.87
ISBN	0.06	0.96
publisher	0.03	0.97
use	0.13	0.87
artist/s	0.03	0.97
owner/s	0.26	0.74
brand name	0.13	0.87
category	0.29	0.80
year of publication	0.03	0.97
geo coord. (lat., long.)	0.19	0.81
type of cuisine	0.06	0.94
population	0.10	0.90
street	0.16	0.84
gps coordinates	0.19	0.81

Table 28: Distinctiveness (eng)

## 4.2 Attributes Relevance

We can combine the two measures considered in the previous section (dominance and distinctiveness) in a single measure, the relevance  $k(c, f)$ , with the following formula:

$$k(c, f) = \phi(c, f) * \psi(f)$$

where  $\psi(f)$  is a logarithmic transformation of the distinctiveness  $\psi_d(f)$

$$\psi(f) = \ln \frac{|C|}{|C(f)|}$$

This measure can be adopted as an estimation of the contribution of an attribute  $f$  to identify a specific category  $c$  and, differently from distinctiveness, may be considered a concept-dependent measure. In other words, if the attribute is used by all (or the majority of) subjects to identify the category (high dominance) and is used only for that specific category (high distinctiveness), the relevance of the attribute for the category is consequently high. This means that the presence of that attribute is highly indicative (that is identifies with high probability) the category considered. For example the attribute “editor” is one of the most frequent attributes for the category *book* in both versions (it results in high values of dominance) and it is reported exclusively in the descriptions of that category (high values of distinctiveness). Combining dominance and distinctiveness we obtain high values of relevance for this attribute when considered respect to the

category *book*. Attributes with high values of relevance are highly informative for entity identification. Continuing our example, consider the query  $q_1$ :<The Lord of the Rings and Allen & Unwin>. If we are able to recognize that “Allen & Unwin” is the name of an editor we can use this information for the entity identification because the presence of the attribute “editor” suggests that the query refers most probably to the book rather than the movie that have the same title “The Lord of the Rings” (namely the same value for the attribute “title”). We calculated the relevance of the attributes listed by the subjects of our experiment, restricting the analysis to attributes reported at least by two subjects, that is we don’t have considered unique attributes. The reason is that unique attributes have the highest distinctiveness but the dominance is minimum so the combination of the two measures can not be significant for our purpose to capture the importance of a given attribute in the distinction of one category from others.

<b>Category</b>	<b>Attributes with <math>k \neq 0</math></b>	<b>Relevance (<math>k</math>)</b>
<i>Politician</i>	partito (party)	35.63
	orientamento politico (political view)	28.67
	nome (name)	22.99
	età (age)	21.02
	cognome (surname)	18.06
	titolo di studio (education)	14.78
	ruolo (role)	14.88
	esperienze (experiences)	12.77
	professione (profession)	11.68
	curriculum	11.68
	carriera (carrier)	7.01
	obiettivo/i (objective/s)	6.87
	fedina penale (criminal record)	6.87
	data di nascita (birth date)	6.56
	luogo di nascita (birth place)	5.47
	figli (children)	4.67
	stato civile (marital status)	4.10
	nazionalità (nationality)	3.71
	provenienza (origin)	3.65
	residenza (residence)	3.28
	domicilio (domicile)	2.74
	posizione sociale (social position)	2.74
	risultati (results)	2.74
	religione (religion)	2.74
	amicizie (friendship)	2.74
	anzianità (seniority)	2.74
	casa (home)	2.74
	impegno (dedication)	2.74
	codice fiscale (tax code)	2.71
	reddito (income)	2.34
	carattere (character)	2.05
	competenze (competence)	2.05
hobbies	2.05	
stato (state)	0.60	

Table 29: Relevance Politician (it)



Category	Attributes with $k \neq 0$	Relevance ( $k$ )
<i>Politician</i>	party	65.78
	name	31.20
	gender	16.42
	position/s	14.60
	age	11.30
	political view	10.30
	background	8.22
	surname	8.21
	biography	6.87
	date elected	6.87
	nationality	6.77
	father's name	5.48
	education	5.42
	role	4.14
	birth place	3.65
	birth date	3.28
	state	2.85
	mother's name	2.74
	responsibility	2.74
	collaborations	2.74
	number of children	2.74
	family	2.74
	passport number	2.34
	occupation	2.34
	qualifications	2.34
	social security number	2.34
	achievements	2.34
	experience	2.34
	language/s	1.82
	country	1.76
	title/s	1.45
	email	0.95
	address	0.90
city	0.54	

Table 30: Relevance Politician (eng)

Category	Attributes with $k \neq 0$	Relevance ( $k$ )
<i>Manager</i>	azienda di appartenenza (membership company)	21.93
	nome (name)	21.34
	cognome (surname)	18.06
	esperienze (experiences)	9.12
	titolo di studio (education)	9.85
	ruolo/i (role/s)	8.93
	età (age)	8.66
	abbigliamento (clothing)	6.87
	disponibilità (availability)	6.87
	punti di forza (strength)	6.87
	qualificazione (qualification)	6.87
	data di nascita (birth date)	6.57
	residenza (residence)	6.57
	competenza (competence)	6.14
	attitudini (attitudes)	5.48
	capacità (capacities)	5.48
	lingue parlate (languages)	5.48
	luogo di lavoro (work place)	5.48
	professione (profession)	4.67
	stato civile (marital status)	4.10
	email	4.06
	laurea (degree)	2.74
	anni di servizio (years on duty)	2.74
	amicizie (friendship)	2.74
	aspettative (expectations)	2.74
	conoscenze (knowledge)	2.74
	reddito(income)	2.74
	religione (religion)	2.34
	aspetto fisico (appearance)	2.34
	figli (children)	2.34
	dati anagrafici (personal details)	2.05
	carattere (character/nature)	2.05
	specializzazione (specialization)	1.82
	progetti (projects)	1.64
altezza (height)	1.49	
codice fiscale (tax code)	1.35	
nazionalità (nationality)	1.24	
attività (activity)	0.79	
stato (state)	0.60	
stipendio (salary)	0.30	

Table 31: Relevance Manager (it)

Category	Attributes with $k \neq 0$	Relevance ( $k$ )
<i>Manager</i>	name	26.27
	experience/s	11.68
	reports	10.30
	department/s	9.12
	occupation	7.01
	budget	6.87
	division/s	6.87
	age	6.08
	responsibility	5.48
	position	5.47
	company	5.47
	gender	4.93
	surname	4.93
	qualifications	4.67
	education	4.06
	organization	3.28
	role	3.11
	phone number	2.90
	title/s	2.90
	email	2.85
	ethnicity	2.74
	intelligent	2.74
	physical characteristics	2.74
	skills	2.74
	supervises	2.74
	affiliation	2.34
	passport number	2.34
	birthday	1.64
	nationality	1.35
	height	1.35
	weight	1.24
	contact	1.04
	address	0.90
	web page	0.60
country	0.18	

Table 32: Relevance Manager (eng)

Category	Attributes with $k \neq 0$	Relevance ( $k$ )
<i>Professor</i>	materia di insegnamento (teaching matter)	54.94
	nome (name)	21.34
	istituzione in cui insegna (institution where teaches)	20.60
	pubblicazioni (publications)	17.17
	cognome (surname)	13.14
	età (age)	11.13
	laurea (degree)	8.22
	corsi (courses)	7.01
	area di ricerca (research area)	6.87
	titolo di studio (education)	6.57
	ruolo/i (role/s)	5.95
	dipartimento (department)	5.48
	anni di servizio (years on duty)	5.48
	specializzazione (specialization)	5.47
	data di nascita (birth date)	4.93
	codice fiscale (tax code)	4.06
	esperienze (experiences)	3.65
	luogo di nascita (birth place)	3.65
	caratteristiche fisiche (appearance)	2.05
	curriculum	2.34
	email	2.71
	anzianità (seniority)	2.74
	fotografia (picture)	2.74
	luogo di lavoro (work place)	2.74
	aspetto fisico (physical appearance)	2.34
	numero di telefono (phone number)	2.26
	competenze (competence)	2.05
	provenienza (origin)	1.82
	livello (level)	1.82
	progetti (projects)	1.64
	residenza (residence)	1.64
	interessi (interests)	1.64
	sito web (web site)	1.04
luogo (place)	0.44	

Table 33: Relevance Professor (it)

Category	Attributes with $k \neq 0$	Relevance ( $k$ )
<i>Professor</i>	university	37.77
	name	34.48
	publication	19.19
	research area	17.17
	department	14.60
	education	9.48
	specialization	9.34
	surname	8.21
	experiences	7.01
	employer	6.87
	school	5.48
	grands	5.48
	middle name	5.48
	courses	5.48
	email	4.75
	faculty/ies	4.67
	affiliations	4.67
	age	4.35
	gender	3.28
	phone number	2.90
	background	2.74
	eye color	2.74
	collaborators	2.74
	intelligent	2.74
	skills	2.74
	supervises	2.74
	web page url	2.40
	qualifications	2.34
	social security number	2.34
	achievements	2.34
	awards	2.05
	birth place	1.82
	position/s held	1.82
	organization	1.64
	birth date	1.64
	title	1.45
	height	1.35
	weight	1.24
	contact	1.04
	address	0.60
city	0.54	
country	0.35	
location	0.22	

Table 34: Relevance Professor (eng)

Category	Attributes with $k \neq 0$	Relevance ( $k$ )
<i>Sportsperson</i>	specialità sportiva (sport specialty)	54.82
	nome (name)	22.99
	età (age)	17.31
	cognome (surname)	11.50
	data di nascita (birth date)	9.85
	livello (level)	10.95
	peso (weight)	7.30
	categoria agonistica (agonistic category)	6.87
	malattie ereditarie (genetic diseases)	6.87
	miglior prestazione (best performance)	6.87
	partecipazione a gare (race participation)	6.87
	tempo dedicato (time dedicated)	6.87
	se fumatore (if smoker)	6.87
	squadra (team)	6.87
	disciplina (discipline)	6.87
	altezza (height)	5.95
	caratteristiche (characteristics)	5.47
	risultati (results)	5.48
	carriera (career)	4.67
	riconoscimenti (award)	4.67
	dati anagrafici (personal details)	4.10
	luogo di nascita (place birth)	3.65
	residenza (residence)	3.28
	ruolo/i (role/s)	2.98
	azienda di appartenenza (membership company)	2.74
	attitudini (attitudes)	2.74
	premi (prizes)	2.74
	vita privata (private )	2.74
	cellulare (mobile phone number)	2.74
	impegno ()	2.74
	qualificazione (qualification)	2.74
	nazionalità (nationality)	2.47
	attività (activity)	2.38
	corsi (courses)	2.34
	numero di telefono (phone number)	2.26
	caratteristiche fisiche (physical appearance)	2.05
	hobbies	2.05
	progetti (projects)	1.64
	email	1.35
	indirizzo (address)	1.20
luogo (place)	0.44	

Table 35: Relevance Sportsperson (it)

<b>Category</b>	<b>Attributes with <math>k \neq 0</math></b>	<b>Relevance (<math>k</math>)</b>
<i>Sportsperson</i>	type of sport	49.34
	name	31.20
	team	17.17
	birth date	11.50
	gender	14.78
	age	8.69
	status professional	6.87
	wins	6.87
	health history	6.87
	nationality	6.77
	awards	6.14
	weight	4.95
	position/s held	3.65
	birth place	3.65
	surname	3.28
	marital status	2.74
	grants	2.74
	alias	2.74
	middle name	2.74
	skin color	2.74
	height	2.71
	specialization	2.34
	achievements	2.34
	picture	1.49
	education	1.35
	mail	0.95
	title/s	0.73
	phone number	0.73
	web page	0.60
	size	0.54
address	0.30	
country	0.18	

Table 36: Relevance Sportsperson (eng)

Category	Attributes with $k \neq 0$	Relevance ( $k$ )
<i>Actor</i>	films	34.34
	nome (name)	24.63
	esperienze (experiences)	25.54
	età (age)	19.79
	cognome (surname)	16.42
	nazionalità (nationality)	13.60
	data di nascita (birth date)	11.50
	titolo di studio (education)	11.50
	cenni biografici (biography)	10.30
	caratteristiche fisiche (physical appearance)	8.19
	carriera (carrier)	7.01
	luogo di nascita (place birth)	7.30
	miglior ruolo interpretato (best part performed)	6.8
	curiosità (curiosity)	6.87
	premi (prizes)	5.48
	vita privata (privacy)	5.48
	codice fiscale (tax code)	2.71
	orientamento politico (political view)	2.05
	altezza (height)	1.49
	carattere (character/nature)	4.10
	stato civile (marital status)	4.10
	colore dei capelli (hair color)	5.48
	famiglia (family)	5.48
	specializzazione (specialization)	3.65
	progetti (projects)	3.28
	ruolo/i (role/s)	2.98
	lingue parlate (languages spoken)	2.74
	capacità (capacity)	2.74
	cellulare (mobile phone number)	2.74
	colore degli occhi (eyes color)	2.74
	reddito (income)	2.34
	curriculum	2.34
	competenze (competence)	2.05
	hobbies	2.05
	dati anagrafici (personal details)	2.05
	provenienza (origin)	1.82
	residenza (residence)	1.64
	interessi (interests)	1.64
	attività (activity)	1.59
	email	1.35
numero di telefono (phone number)	1.13	
indirizzo (address)	0.60	

Table 37: Relevance Actor (it)



Category	Attributes with $k \neq 0$	Relevance ( $k$ )
<i>Actor</i>	name	26.27
	films	19.19
	birth date	11.50
	gender	9.85
	awards	6.14
	surname	4.93
	role	4.14
	height	4.06
	birth place	3.65
	language	3.65
	age	3.48
	facial feature	2.74
	family	2.74
	origin	2.74
	physical characteristics	2.74
	nationality	2.71
	activity	2.34
	character/s	2.34
	category	2.47
	weight	2.47
	organization	1.64
	picture	1.49
	education	1.35
	color	0.95
	email	0.95
	country	0.88
	phone number	0.73
	size	0.54
	address	0.30

Table 38: Relevance Actor (eng)

Category	Attributes with $k \neq 0$	Relevance ( $k$ )
<i>Person</i>	nome (name)	32.84
	cognome (surname)	27.92
	luogo di nascita (birth place)	18.25
	professione (occupation)	18.68
	data di nascita (birth date)	16.42
	colore degli occhi (eyes color)	13.70
	età (age)	12.37
	carattere (character)	12.29
	residenza (residence)	11.50
	codice fiscale (tax code)	10.84
	colore dei capelli (hair color)	10.96
	segni particolari (particularity)	10.30
	altezza (height)	10.42
	peso (weight)	10.95
	colore della pelle (skin color)	6.87
	impronte digitali (fingerprint)	6.87
	dati anagrafici (personal details)	6.14
	hobbies	6.14
	posizione sociale (social position)	5.48
	domicilio (domicile)	5.48
	posizione sociale (social position)	5.48
	aspetto fisico (physical appearance)	4.67
	stato civile (marital status)	4.10
	nazionalità (nationality)	3.71
	famiglia (family)	2.74
	aspettative (expectancy)	2.74
	fotografia (picture)	2.74
	religione (religion)	2.74
	figli (children)	2.34
	caratteristiche fisiche (physical characteristics)	2.05
	esperienze (experiences)	1.82
	provenienza (origin)	1.82
	interessi (interests)	1.64
	titolo di studio (education)	1.64
sito web (web site)	1.04	
attività (activity)	0.79	
stato (state)	0.60	
indirizzo (address)	0.60	

Table 39: Relevance Person (it)

Category	Attributes with $k \neq 0$	Relevance ( $k$ )
<i>Person</i>	name	31.20
	gender	22.99
	birth date	18.06
	occupation	14.01
	religion	13.74
	marital status	10.96
	mother name	10.96
	parents name	10.30
	hair color	10.30
	surname	9.85
	birth place	9.12
	age	8.69
	father's name	8.22
	eyes color	8.22
	nationality	8.13
	height	8.13
	education	8.13
	id	6.87
	residence	6.87
	language/s	5.47
	email	4.75
	passport number	4.67
	social security number	4.67
	picture	4.46
	position	3.65
	films	2.74
	school	2.74
	alias	2.74
	ethnicity	2.74
	facial features	2.74
	number of children	2.74
	origin	2.74
	skin color	2.74
	weight	2.47
	activity	2.34
	affiliation/s	2.34
	character	2.34
	phone number	2.18
	awards	2.05
	organization	1.64
	contact	1.04
role	1.04	
color	0.95	
address	0.90	
title/s	0.73	
web page url	0.60	

location	0.43
country	0.18

Table 40: Relevance Person (eng)

Category	Attributes with $k \neq 0$	Relevance ( $k$ )
<i>Company</i>	nome (name)	36.12
	numero di dipendenti (number of employees)	14.78
	fatturato (turnover)	12.29
	capitale sociale ( share capital)	10.96
	produzione (output)	10.30
	struttura (structure)	8.22
	prodotti (products)	6.87
	localizzazione (localization)	6.87
	soci (partners)	6.87
	meccanismo di controllo (control mechanism)	6.87
	storia (history)	6.14
	partita iva (VAT number)	5.95
	settore (sector)	5.66
	strategie ( strategies)	5.48
	processi decisionali (decisional processes)	5.48
	funzioni (functions)	5.48
	luogo (place)	4.82
	organigramma (organization chart)	4.10
	responsabile (responsible)	4.10
	sede (main office)	4.06
	proprietario (owner)	4.06
	regione (region)	3.65
	numero di telefono (phone number)	3.39
	data di fondazione (date of foundation)	3.28
	ruolo (role)	2.98
	persone (people)	2.74
	vision	2.74
	azionisti (shareholders)	2.74
	campo occupazionale (occupational field)	2.74
	capo (boss)	2.74
	grandezza (size)	2.74
	idee (ideas)	2.74
	impiegati (employs)	2.74
	nome amministratore delegato (managing director name)	2.74
	numero di fax (fax number)	2.74
	responsabilità (responsability)	2.74
	codice fiscale (tax code)	2.71
	ente organizzativo (organizational company )	2.47
	mission	2.34
	presidente (president)	2.34

professione (occupation)	2.34
riconoscimenti (awards)	2.34
membri (members)	2.34
sede legale ( office)	2.34
logo	2.34
servizi (services)	2.26
sito web (web site)	2.07
ambito (area)	2.05
finalità (goals)	2.05
dimensione/i (dimension/s)	1.98
specializzazione (specialization)	1.82
indirizzo (address)	1.80
attività (activity)	1.59
scopo/i (objective/s)	1.45
categoria di appartenenza (category)	1.35
email	1.35
stato (state)	1.20
città (city)	0.60

---

Table 41: Relevance Company (it)

<b>Category</b>	<b>Attributes with <math>k \neq 0</math></b>	<b>Relevance (<math>k</math>)</b>
<i>Company</i>	name	24.63
	ceo name	8.22
	business type	8.19
	profits	7.01
	revenue	6.87
	number of employees	5.95
	logo	5.48
	board members	5.48
	product/s	4.67
	industry/ies	4.67
	head quarters	4.10
	owner/s	4.06
	contact	3.11
	web page url	2.40
	financial data	2.74
	head quarters location	2.74
	history	2.74
	partners	2.74
	symbol	2.74
	annual turnover	2.74
	director/s	2.34
	chairman	2.34
	phone number	2.18
	address	2.09
	brand name	2.05
	business registration number	2.05
	sector	2.05
	state	1.90
	date of foundation	1.82
	company	1.82
	location	1.51
	service/s	1.49
	nationality	1.35
	city	1.09
	size	1.09
	role	1.04
country	1.06	
title	0.73	
type	0.34	

Table 42: Relevance Company (eng)

Category	Attributes with $k \neq 0$	Relevance ( $k$ )
<i>Association</i>	nome (name)	27.91
	associati (members)	27.47
	scopo/i (objective/s)	11.61
	numero iscritti (number of members registered)	7.01
	funzioni (functions)	6.87
	settore (sector)	6.79
	sede (main office)	6.77
	data di fondazione (date of foundation)	4.93
	presidente (president)	4.67
	attività (activity)	3.18
	cateoria (category)	2.98
	processi decisionali (decision processes)	2.74
	responsabile (responsible)	2.74
	contatto/i (contact/s)	2.74
	rappresentante legale (legal representative)	2.74
	soci (partners)	2.74
	numero componenti (number of members)	2.34
	sede legale (registered office)	2.34
	logo	2.34
	professione (occupation)	2.34
	luogo (location)	2.19
	ambito (area)	2.05
	fatturato (turnover)	2.05
	orientamento politico (political view)	2.05
	finalità (goals)	2.05
	regione (region)	2.05
	prgetti (projects)	1.64
	numero di dipendenti (number of employees)	1.64
	partita iva (VAT number)	1.49
	codice fiscale (tax code)	1.35
	campo d'azione (area of influence)	1.35
	ente organizzatore (organizational company)	1.24
	indirizzo (address)	1.20
	servizi (services)	1.13
	numero di telefono (phone number)	1.13
	sito web (web site)	1.04
dimensione/i (dimension/s)	0.66	

Table 43: Relevance Association (it)

Category	Attributes with $k \neq 0$	Relevance ( $k$ )
<i>Association</i>	name	21.34
	objective/s	16.42
	members	11.68
	activity	11.68
	date of foundation	9.12
	leader/s	8.22
	international	6.87
	committee	6.87
	membership	6.87
	domain	6.87
	logo	5.48
	number of members	4.67
	web page	3.60
	funding	4.67
	services	2.98
	number of employees	2.98
	ceo name	2.74
	board members	2.74
	achievements	2.74
	annual turnover	2.74
	field	2.74
	financial data	2.74
	head quarters	2.74
	history	2.74
	managers	2.74
	representation	2.74
	short name	2.74
	start date	2.74
	number of students	2.74
	symbol	2.74
	chairman	2.34
	legal form	2.34
	scale	2.34
	head	2.34
	profit/s	2.34
	size	2.17
	role	2.07
	contact/s	2.07
	type	2.06
	business registration number	2.05
	head quarters	2.05
	business type	2.05
sector	2.05	
email	1.90	
location	1.72	
city	1.63	
address	1.49	
nationality	1.35	
participants	1.24	
phone number 64	0.73	
title/s	0.73	
country	0.53	

Table 44: Relevance Association (eng)



Category	Attributes with $k \neq 0$	Relevance ( $k$ )
<i>University</i>	facoltà (faculties)	30.91
	nome (name)	22.99
	numero di studenti (number of students)	17.17
	corsi (courses)	16.35
	docenti (professors)	13.74
	corsi di laurea (degree course)	10.30
	dipartimento/i (department/s)	8.22
	struttura (structure)	8.22
	numero iscritti (number of students)	7.01
	associazioni (associations)	6.87
	metodologia didattica (teaching approach)	6.87
	modalità di accesso (admission way)	6.87
	offerta formativa (educational offers)	6.87
	campus	6.87
	master	6.87
	luogo (location)	6.57
	servizi (services)	5.66
	titolo di studio (qualifications)	4.93
	storia (history)	4.10
	sede (main office)	4.06
	città (city)	3.60
	missione (mission)	2.74
	mission	2.34
	sito web (web site)	2.07
	organigramma (organization chart)	2.05
	personale (staff)	2.05
	indirizzo (address)	1.80
	progetti (projects)	1.64
	data di fondazione (date of foundation)	1.64
	ente organizzativo (organization responsible)	1.24
	scopo/i (goal/s)	1.45
	contatto/i (contact/s)	1.49
	categoria di appartenenza (category)	1.35
	numero di telefono (phone number)	1.13
descrizione (description)	0.87	
attività (activity)	0.79	
dimensione/i (dimension/s)	0.66	
stato (state)	0.60	

Table 45: Relevance University (it)

Category	Attributes with $k \neq 0$	Relevance ( $k$ )
<i>University</i>	name	26.27
	number of students	19.19
	faculty/ies	16.35
	courses	13.70
	department/s	9.12
	international programs	6.87
	number of majors	6.87
	number of professors	6.87
	rank	6.87
	research programs	6.87
	admission	6.87
	staff	6.87
	publication	5.48
	date of foundation	5.47
	state	4.75
	province	4.46
	objective/s	3.28
	city	3.81
	acronym	2.74
	start date	2.74
	location	2.58
	address	2.39
	funding	2.34
	scale	2.34
	education	1.35
	size	1.09
	country	1.06
	age	0.87
	phone number	0.73
	web page	0.60
type	0.69	

Table 46: Relevance University(eng)

<b>Category</b>	<b>Attributes with <math>k \neq 0</math></b>	<b>Relevance (<math>k</math>)</b>
<i>Government</i>	orientamento politico (political view)	16.38
	durata (duration)	13.74
	partito/i (party/s)	10.96
	ministeri (ministries)	10.30
	ministri (ministers)	10.30
	premier	10.30
	nazione (nation)	10.42
	leggi (laws)	6.87
	ordinamento (ordination)	6.87
	capo del governo (premier)	6.87
	formazione (formation)	6.87
	persone (people)	5.48
	data di fondazione (date of foundation)	4.93
	membri (members)	4.67
	numero componenti (number of members)	4.67
	ruolo (role)	2.98
	bilancio (budget)	2.74
	capo (boss)	2.74
	idee (ideas)	2.74
	nazionalità (nationality)	2.47
	stato (state)	2.40
	storia (history)	2.05
	ambito (area)	2.05
	responsabile (responsible)	2.05
	nome (name)	1.64
	attività (activity)	0.79
	scopo/i (objective/s)	0.73

Table 47: Relevance Government(it)

Category	Attributes with $k \neq 0$	Relevance ( $k$ )
<i>Government</i>	name	11.49
	head	9.34
	members	9.34
	party	8.22
	leaders	8.22
	level	6.87
	characteristics	5.48
	year	2.71
	country	2.64
	number of members	2.34
	legal form	2.34
	department	1.82
	objectives	1.64
	province	1.49
	education	1.35
	web page	1.20
	role	1.04
	type	1.03
	state	0.95
	email	0.95
phone number	0.73	
size	0.54	
address	0.30	
location	0.22	

Table 48: Relevance Government (eng)

Category	Attributes with $k \neq 0$	Relevance ( $k$ )
<i>Agency</i>	nome (name)	26.27
	numero di dipendenti (number of employees)	8.21
	clientela (clients)	6.87
	settore (sector)	4.53
	scopo/i (objective/s)	4.36
	personale (staff)	4.10
	indirizzo (address)	3.60
	servizi (services)	3.39
	luogo (location)	3.07
	proprietario (owner)	4.06
	partita iva (VAT number)	2.98
	strategie (strategies)	2.74
	ambito di lavoro (area )	2.74
	campo di azione ( influence area)	2.74
	campo occupazionale (professional field)	2.74
	impiegati (employees)	2.74
	numero di fax (fax number)	2.74
	utilità (utility)	2.74
	forma giuridica (legal status )	2.74
	sede (main office)	2.71
	logo	2.34
	professione (profession)	2.34
	responsabile (responsible)	2.05
	organigramma (organizational chart)	2.05
	fatturato (turnover)	2.05
	ambito (area)	2.05
	email	1.35
	dimensione/i (dimension/s)	1.32
	ente organizzatore (organizational company)	1.24
	numero di telefono (phone number)	1.13
	descrizione (description)	0.87
	città (city)	0.60
	stato (state)	0.60

Table 49: Relevance Agency (it)

Category	Attributes with $k \neq 0$	Relevance ( $k$ )
<i>Agency</i>	name	21.34
	number of employees	7.44
	president	6.87
	specialization	4.67
	profit/s	4.67
	product/s	4.67
	funding	4.67
	sector	4.10
	business type	4.10
	services	2.98
	acronym	2.74
	managers	2.74
	partners	2.74
	short name	2.74
	industry/ies	2.34
	director/s	2.34
	address	2.09
	brand name	2.05
	business registration number	2.05
	head quarters	2.05
	date of foundation	1.82
	department/s	1.82
	type	1.71
	owner/s	1.35
	nationality	1.35
	participants	1.24
	contact	1.04
	role	1.04
	phone number	0.73
	location	0.65
	web page	0.60
	size	0.54
	country	0.18

Table 50: Relevance Agency (eng)

<b>Category</b>	<b>Attributes with <math>k \neq 0</math></b>	<b>Relevance (<math>k</math>)</b>
<i>Organization</i>	nome (name)	27.91
	scopo/i (objective/s)	12.34
	membri (members)	11.68
	settore (sector)	9.05
	data di fondazione (date of foundation)	8.21
	sede (main office)	8.13
	mission	7.01
	etica (ethic)	6.87
	organigramma (organizational chart)	6.14
	forma giuridica (legal status)	5.48
	vision	5.48
	numero dipendenti (number of employees)	4.93
	sede legale (registered office)	4.67
	numero componenti (number of members)	4.67
	nazione (nation)	4.46
	attività (activity)	3.18
	luogo (location)	3.07
	partita iva (VAT number)	2.98
	capitale sociale (share capital)	2.74
	ambito di lavoro (professional area)	2.74
	azionisti (shareholders)	2.74
	bilancio (budget)	2.74
	grandezza (size)	2.74
	missione (mission)	2.74
	prodotti (products)	2.74
	rappresentante legale (legal representative)	2.74
	utilità (utility)	2.74
	dimensione/i (dimension/s)	2.65
	nazionalità (nationality)	2.47
	presidente (president)	2.34
	riconoscimenti (honors)	2.34
	numero iscritti (number of members)	2.34
	fatturato (turnover)	2.05
	specializzazione (specialization)	1.82
	regione (region)	1.82
	codice fiscale (tax code)	1.35
	proprietario (owner)	1.35
	stato (state)	1.20
	città (city)	1.20
	sito web (web site)	1.04
descrizione	0.87	
indirizzo (address)	0.60	

Table 51: Relevance Organization (it)

Category	Attributes with $k \neq 0$	Relevance ( $k$ )
<i>Organization</i>	name	21.34
	business type	6.14
	objective/s	4.93
	character/s	4.67
	head	4.67
	product/s	4.67
	business registration number	4.10
	date of foundation	1.82
	number of employees	2.98
	achievements	2.74
	field	2.74
	representation	2.74
	industry/ies	2.34
	faculty/ies	2.34
	members	2.34
	number of members	2.34
	legal form	2.34
	scale	2.34
	role	2.07
	type	2.06
	head quarters	2.05
	sector	2.05
	company	1.82
	age	1.74
	location	1.29
	contact	1.04
	email	0.95
	address	0.90
	phone number	0.73
	city	0.54
country	0.53	

Table 52: Relevance Organization (eng)



<b>Category</b>	<b>Attributes with <math>k \neq 0</math></b>	<b>Relevance (<math>k</math>)</b>
<i>Conference</i>	argomento (topic)	25.54
	relatori (speakers)	16.45
	partecipanti (participants)	16.42
	data (date)	11.13
	necessità (needs)	10.30
	anno (year)	10.30
	durata (duration)	9.85
	titolo (title)	8.21
	luogo (location)	7.89
	ora (time)	5.95
	settore ( sector)	4.53
	orario di inizio (start time)	4.67
	ente organizzativo (organizers)	4.95
	destinatari (recipients)	3.65
	nome dell'evento (name of the event)	3.64
	contatto/i (contact/s)	2.98
	promotori (promoters)	2.74
	agenda	2.74
	giorno (day)	2.05
	moderatore (moderator)	2.34
	prezzo di ingresso (price of entry)	2.05
	programma (program)	2.05
	autore (author)	2.05
	finalità (goals)	2.05
	quota di partecipazione (share)	1.64
	sponsor/s	1.64
descrizione (description)	0.87	

Table 53: Relevance Conference (it)

Category	Attributes with $k \neq 0$	Relevance ( $k$ )
<i>Conference</i>	name	22.32
	organizers	12.77
	date	12.37
	chair/s	10.30
	sessions	10.30
	topic/s	10.24
	sponsor/s	8.19
	start date	7.01
	fees	6.87
	acceptance	6.87
	year	6.77
	end date	6.57
	participants	6.18
	subject	8.21
	purpose/s	5.42
	location	4.73
	audience	4.67
	kind	4.10
	organization	3.28
	time	2.98
	duration	2.98
	cost/s	2.74
	effect	2.74
	length	2.74
	related events	2.74
	series	2.74
	who	2.34
	what	2.34
	title	2.18
	objective/s	1.64
	start time	1.64
	web page url	1.20
type	0.69	
address	0.30	
country	0.18	

Table 54: Relevance Conference (eng)

Category	Attributes with $k \neq 0$	Relevance ( $k$ )
<i>Meeting</i>	argomento (topic)	20.07
	ora (time)	17.86
	partecipanti (participants)	16.42
	data (date)	11.13
	luogo (location)	7.01
	relatore (speaker)	5.48
	destinatari (recipients)	3.65
	settore (area)	3.39
	durata (duration)	3.28
	giorno (day)	2.05
	ora di inizio (start time)	2.05
	modalità di partecipazione (ways of participation)	2.74
	agenda	2.74
	modalità di svolgimento (carrying out )	2.74
	tematica (basic theme)	2.74
	motivo (cause)	2.34
	indirizzo (address)	1.80
	interesse (interest)	1.64
città (city)	1.20	
stato (state)	0.60	

Table 55: Relevance Meeting (it)

Category	Attributes with $k \neq 0$	Relevance ( $k$ )
<i>Meeting</i>	time	29.76
	date	19.79
	topic/s	18.43
	participants	16.08
	agenda	13.74
	minutes	10.30
	subject	9.85
	name	7.44
	discussions	6.87
	ways of communication	6.87
	presentation	6.87
	room number	6.87
	speaker/s	5.48
	organizer/s	5.47
	location	4.73
	chairman	4.67
	duration	4.46
	purpose/s	4.06
	content	2.74
	schedule	2.74
	stage	2.74
	subject	2.74
	what	2.34
	host	2.34
	who	2.34
	title	2.18
	type	2.06
	email	1.90
	start time	1.64
	end date	1.64
	objective/s	1.64
	description	1.04
	contact	1.04
	address	0.60
web page	0.60	
city	0.54	
country	0.18	

Table 56: Relevance Meeting (eng)

Category	Attributes with $k \neq 0$	Relevance ( $k$ )
<i>Exhibition</i>	argomento (topic)	21.89
	durata (duration)	11.50
	artisti partecipanti (artists)	10.96
	espositori (exhibitors)	10.30
	titolo (title)	9.85
	artista (artist)	9.34
	data (date)	8.66
	luogo (location)	7.01
	nome (name)	7.29
	prezzo di entrata ( price of entry)	6.14
	ora (time)	4.46
	sede (site)	4.06
	ente organizzativo (organizers)	3.71
	prezzo (price)	2.85
	data di inizio (start date)	2.74
	necessità (needs)	2.34
	target	2.34
	data fine (end date)	2.74
	numero (number)	2.74
	sito web (web site)	2.07
	soggetto (subject)	2.05
	autore (author)	1.64
	partecipanti (participants)	1.64
	interessi (interests)	1.64
	indirizzo (address)	1.20
	settore (sector)	1.13
	servizi (services)	1.13
	descrizione (description)	0.87

Table 57: Relevance Exhibition (it)

Category	Attributes with $k \neq 0$	Relevance ( $k$ )
<i>Exhibition</i>	name	11.90
	time	8.93
	date	7.42
	start date	7.01
	end date	4.93
	subject	3.28
	duration	2.98
	speakers	2.74
	target audience	2.74
	content	2.74
	theme	2.74
	year	2.71
	location	2.58
	impact	2.34
	description	2.07
	topic/s	2.05
	sponsor/s	2.05
	participants	1.24
	role	1.04
	contact	1.04
title	0.73	
web page	0.60	
city	0.54	
type	0.34	
country	0.18	

Table 58: Relevance Exhibition (eng)

Category	Attributes with $k \neq 0$	Relevance ( $k$ )
<i>Show</i>	data (date)	19.79
	attori (actors)	18.68
	durata (duration)	18.06
	ora (time)	16.37
	titolo (title)	14.78
	luogo (location)	11.83
	regia (direction)	10.30
	regista (director)	10.30
	partecipanti (participants)	9.85
	prezzo (prize)	9.49
	artisti (artists)	7.01
	ideatore (creator)	6.87
	intrattenimento (entertainments)	6.87
	musica (music)	6.87
	numero di spettatori (spectators)	6.87
	scenografia (scenogrphy )	6.87
	cast	6.87
	ambientazione (setting)	6.87
	artisti partecipanti (artists)	5.48
	nome (name)	3.64
	autore (authors)	3.28
	prezzi di ingresso (prize of entry)	2.05
	giorno (day)	2.05
	programma (program)	2.05
	sponsor/s	2.05
	antefatti(prior events)	2.74
	staff	2.74
	descrizione (description)	2.61
	ente organizzatore (organizers)	2.47
	persone coinvolte (people involved)	2.34
	target (director)	2.34
	collegamenti (connections)	1.82
	interessi(interests)	1.64
	contatto/i (contact/s)	1.49
settore (sector)	1.13	
servizi (services)	1.13	

Table 59: Relevance Show (it)

Category	Attributes with $k \neq 0$	Relevance ( $k$ )
<i>Show</i>	actors	17.17
	name	13.39
	producer/s	10.30
	time	7.44
	director/s	7.01
	performer/s	6.87
	protagonist	6.87
	target audience	5.48
	date	4.95
	end date	4.93
	sponsor/s	4.10
	participants	3.71
	duration	1.49
	organizer/s	1.82
	start time	3.28
	subject	3.28
	length	2.74
	schedule	2.74
	stage	2.74
	subject	2.74
	theme	2.74
	start date	2.34
	audience	2.34
	host	2.34
	location	1.94
	state	1.90
	city	1.63
	title	1.45
	type	1.37
	color	0.95
	address	0.60
	country	0.35

Table 60: Relevance Show (eng)



Category	Attributes with $k \neq 0$	Relevance ( $k$ )
<i>Incident</i>	persone coinvolte (people involved)	25.69
	entità coinvolte (entities involved)	24.04
	danni (damages)	13.74
	dinamica (dynamics)	13.74
	veicoli coinvolti (vehicles involved)	10.30
	ora (time)	11.90
	data (date)	8.66
	conseguenze (consequences)	8.22
	luogo (location)	7.89
	fattori causali (causal factors)	6.87
	gravità (gravity)	6.87
	attori (actors)	4.67
	giorno (day)	2.05
	motivo (cause)	2.34
	antefatti (prior events)	2.74
	entità (extent)	2.74
	mese (month)	2.74
	modalità di svolgimento (course of events)	2.74
	via (street)	1.64
	anno (year)	1.64
descrizione (description)	0.87	
città (city)	0.60	
stato (state)	0.60	

Table 61: Relevance Incident (it)

Category	Attributes with $k \neq 0$	Relevance ( $k$ )
<i>Incident</i>	time	20.83
	date	11.13
	people involved	10.96
	participants	7.42
	causes	7.01
	severity	6.87
	type of incident	6.87
	consequences	6.87
	duration	4.46
	organization	3.28
	start time	3.28
	outcome	2.74
	effect	2.74
	location	2.58
	results	2.34
	impact	2.34
	description	2.07
	end date	1.64
	name	1.48
	type	1.37
year	1.35	
purpose/s	1.35	
role	1.04	
title	0.73	
city	0.54	
country	0.35	

Table 62: Relevance Incident (eng)

Category	Attributes with $k \neq 0$	Relevance ( $k$ )
<i>Event</i>	data (date)	18.55
	partecipanti (participants)	13.14
	nome (name)	12.77
	durata (duration)	8.21
	ora (time)	7.44
	motivazioni (motivations)	6.87
	invitati (attendants)	6.87
	luogo (location)	6.57
	ente organizzativo (organizers)	6.18
	contatto/i (contact/s)	5.95
	giorno (day)	4.10
	programma (program)	4.10
	titolo (title)	3.28
	prezzo (price)	2.85
	conseguenze (consequences)	2.74
	mese (month)	2.74
	ospiti (guests)	2.74
	sede (place)	2.71
	descrizione(description)	2.61
	attività (activity)	2.38
	quota di partecipazione (participation price)	2.34
	motivo (cause)	2.34
	necessità (needs)	2.34
	target	2.34
	persone coinvolte (people involved)	2.34
	attore (actors)	2.34
	sito web (web site)	2.07
	soggetto (subject)	2.05
	orario di inizio (start time)	2.05
	sponsor/s	2.05
	collegamenti (connections)	1.82
	argomento (topic)	1.82
	anno (year)	1.64
	nazione (nation)	1.49
	indirizzo (address)	1.20
	settore (sector)	1.13
città (city)	0.60	

Table 63: Relevance Event (it)

Category	Attributes with $k \neq 0$	Relevance ( $k$ )
<i>Event</i>	time	17.86
	date	17.31
	name	10.41
	participants	9.89
	repetition	6.87
	purpose/s	6.77
	outcome	5.48
	organizers	5.47
	what	4.67
	location	4.52
	duration	4.46
	description	3.11
	people involved	2.74
	competitors	2.74
	prize/s	2.74
	audience	2.34
	causes	2.34
	results	2.34
	host	2.34
	impact	2.34
	contact	2.07
	type	2.06
	sponsor/s	2.05
	topic/s	2.05
	start time	1.64
	end date	1.64
	color	0.95
	web page url	0.60
	city	0.54

Table 64: Relevance Event (eng)

Category	Attributes with $k \neq 0$	Relevance ( $k$ )
<i>Sports event</i>	specialità sportiva (type of sport)	27.41
	nome (name)	14.59
	data (date)	13.60
	luogo (location)	7.01
	ora (time)	5.95
	data di inizio (start date)	5.48
	modalità di partecipazione (participation ways)	5.48
	partecipanti (participants)	6.57
	sponsor/s	4.10
	livello (level)	3.65
	destinatari (recipients)	3.65
	ente organizzativo (organizers)	3.71
	durata (duration)	3.28
	data di fine (end date)	2.74
	entità (extent)	2.74
	numero (number)	2.74
	ospite/i (guests)	2.74
	staff	2.74
	quota di partecipazione (participation price)	2.34
	prezzo di ingresso (entry price)	2.05
	programma (programs)	2.05
	orario di inizio (start time)	2.05
	finalità (finality)	2.05
	via (street)	1.64
	nazione (nation)	1.49
	contatto/i (contact/s)	1.49
	sito web (web site)	1.04
	città (city)	0.60

Table 65: Relevance Sports Event (it)

Category	Attributes with $k \neq 0$	Relevance ( $k$ )
<i>Sports Event</i>	type of sport	30.15
	stadium	10.30
	date	12.37
	time	7.44
	winners	6.87
	game	6.87
	players	6.87
	duration	5.95
	name	5.95
	participants	4.95
	results	4.67
	location	3.87
	organizer/s	3.65
	competition	2.74
	cost/s	2.74
	prize/s	2.74
	related events	2.74
	series	2.74
	causes	2.34
	contact	2.07
	kind	2.05
	start time	1.64
	type	1.03
state	0.95	
city	0.54	
address	0.30	
country	0.18	

Table 66: Relevance Sports Event (eng)

<b>Category</b>	<b>Attributes with <math>k \neq 0</math></b>	<b>Relevance (<math>k</math>)</b>
<i>Product</i>	utilizzo (use)	22.52
	nome (name)	18.06
	prezzo (price)	14.24
	colore (color)	11.13
	marca (brand)	9.34
	caratteristiche (features)	9.12
	peso (weight)	9.12
	qualità (quality)	8.22
	funzione (function)	7.01
	codice (code)	6.87
	componenti (components)	6.87
	luogo di vendita (dealer)	6.87
	scadenza (due date)	6.87
	composizione (composition)	5.48
	forma (shape)	4.93
	produttore (producer)	4.67
	dimensione/i (dimension/s)	4.63
	provenienza (origin)	3.65
	funzionalità (functionality)	2.74
	odore (smell)	2.74
descrizione (description)	2.61	
misure (size)	2.05	
materiale (material)	1.64	
categoria di appartenenza (category)	1.35	
scopo/i (aim/s)	0.73	

Table 67: Relevance Product (it)

Category	Attributes with $k \neq 0$	Relevance ( $k$ )
<i>Product</i>	manufacturer	21.02
	price/s	14.84
	name	14.78
	use	14.33
	warranty	10.30
	availability	6.87
	brand name	6.14
	characteristic/s	5.48
	description	5.18
	image	5.48
	material	3.65
	ISBN	2.74
	related to	2.74
	where available	2.74
	dimension	2.34
	function	2.34
	id	2.34
	author	2.34
	size	2.17
	kind	2.05
	creation date	2.05
	company	1.82
	shape	1.64
	picture	1.49
	title	1.45
	purpose	1.35
	date	1.24
category	1.24	
weight	1.24	
type	1.03	

Table 68: Relevance Product (eng)



<b>Category</b>	<b>Attributes with <math>k \neq 0</math></b>	<b>Relevance (<math>k</math>)</b>
<i>Artwork</i>	autore (author)	37.77
	luogo (location)	24.57
	stile (style)	18.68
	data di creazione (creation date)	17.17
	tecnica ( technique)	17.17
	titolo (title)	14.78
	materiale (material)	14.78
	soggetto (subject)	12.29
	periodo (age)	10.30
	artista (artist)	9.34
	data (date)	9.89
	anno (year)	8.21
	colore (color)	7.42
	significato (meaning)	6.87
	museo (museum)	6.87
	nome (name)	6.56
	epoca (age)	5.48
	dimensione/i (dimension/s)	3.31
	forma (shape)	3.28
	misure (size)	2.05
spessore (thickness)	2.05	
data di fine (end date)	2.74	
proprietario (owner)	2.71	
descrizione (description)	1.74	
categoria di appartenenza (category)	1.35	
città (city)	1.20	

Table 69: Relevance Artwork (it)

Category	Attributes with $k \neq 0$	Relevance ( $k$ )
<i>Artwork</i>	artist/s	30.91
	creation date	18.43
	style	14.01
	material	10.95
	author	7.01
	display location	6.87
	technique	6.87
	title	5.81
	date	4.95
	name	4.92
	creator	2.74
	image	2.74
	subject	1.64
	owner/s	2.71
	price	2.47
	id	2.34
	dimensions	2.34
	width	2.34
	place	2.05
	color	1.90
	type	1.71
	shape	1.64
	size	1.63
	year	1.35
	height	1.35
	category	1.24
	description	1.04
	age	0.87
	location	0.22
	country	0.18

Table 70: Relevance Artwork (eng)

Category	Attributes with $k \neq 0$	Relevance ( $k$ )
<i>Building</i>	numero di piani (number of floors)	37.77
	luogo (location)	32.76
	metratura (mq)	20.60
	altezza (height)	16.37
	anno di costruzione (date of creation)	13.74
	colore (color)	12.37
	utilizzo (use)	12.29
	destinatari (recipients)	10.95
	architetto (architect)	10.30
	larghezza (width)	8.22
	numero di stanze (number of rooms)	5.48
	didascalia (caption)	6.87
	ditta costruttrice ()	6.87
	dimensione/i (dimension/s)	6.61
	materiale (material)	4.93
	numero civico (building/ street number)	4.67
	stile (style)	4.67
	stile architettonico (architectural style)	4.67
	indirizzo (address)	4.21
	posizione (position)	4.10
	cap (zip code)	4.10
	data di fine opera (end date of creation)	2.74
	età (age)	2.47
	superficie (surface)	2.34
	funzione (function)	2.34
	misure (size)	2.05
	spessore (thickness)	2.05
	prezzo (price)	1.90
	peso (weight)	1.82
	nome (name)	1.64
	forma (shape)	1.64
	caratteristiche (characteristics)	1.82
	proprietario (owner)	1.35
	città (city)	1.20
stato (state)	0.60	
scopo (aim)	0.73	

Table 71: Relevance Building (it)

Category	Attributes with $k \neq 0$	Relevance ( $k$ )
<i>Building</i>	architect	20.60
	number of floors	17.17
	height	14.90
	name	13.13
	architectural style	10.30
	floors	10.30
	use	8.19
	style	7.01
	number	7.01
	direction	6.87
	area	6.87
	date of construction	6.87
	owner/s	6.77
	material	5.47
	gps coordinates	4.93
	color	4.75
	width	4.67
	address	4.48
	creation date	4.10
	street	3.65
	geo coordinates	3.28
	location	2.80
	purpose/s	2.71
	usage	2.34
	function	2.34
	place	2.05
	state	1.90
	type	1.71
	shape	1.64
	city	1.63
	size	1.63
	picture	1.49
year	1.35	
country	0.70	

Table 72: Relevance Building (eng)

Category	Attributes with $k \neq 0$	Relevance ( $k$ )
<i>Book</i>	editore (editor)	61.81
	numero di pagine (number of pages)	54.94
	autore (author)	44.34
	titolo (title)	36.13
	anno di pubblicazione (publication date)	30.91
	ISBN	24.04
	edizione (edition)	13.74
	lingua (language)	10.96
	argomento (argument)	10.95
	anno di edizione (edition date)	6.87
	anno di prima edizione (first edition)	6.87
	collana (series)	6.87
	colore della copertina (book cover color)	6.87
	copertina (book cover)	6.87
	traduttore (translator)	6.87
	trama (plot)	6.87
	contenuto (content)	5.48
	materiale $\delta$ material)	4.93
	prezzo (price)	3.80
	nome (name)	/3.28
	fascia d'età (age group)	2.74
	colore (color)	2.47
	descrizione (description)	2.61
	larghezza (width)	2.74
	soggetto (subject)	2.05
	spessore (thickness)	2.05
	anno (year)	1.64
	forma (shape)	1.64
	altezza (height)	1.49
	data (date)	1.24
	dimensione/i (dimension/s)	0.66
	stato (state)	0.60
	città (city)	0.60

Table 73: Relevance Book (it)

Category	Attributes with $k \neq 0$	Relevance ( $k$ )
<i>Book</i>	author/s	46.71
	publisher	44.64
	ISBN	35.63
	year of publication	27.47
	number of pages	24.04
	title	13.79
	edition	10.30
	subject	8.21
	language/s	7.30
	editor	6.87
	year	6.77
	name	4.92
	price	3.71
	unique identity number	2.74
	theme	2.74
	category	2.47
	texture	2.34
	width	2.34
	color	1.90
	shape	1.64
height	1.35	
weight	1.24	
type	0.69	
address	0.30	

Table 74: Book (eng)

Category	Attributes with $k \neq 0$	Relevance ( $k$ )
<i>Article of clothing</i>	taglie (sizes)	72.11
	marca (brand)	35.03
	colore (color)	33.39
	tessuto (fabric)	27.47
	modello (model)	17.17
	prezzo (price)	11.39
	stile (style)	9.34
	materiale (material)	8.21
	per che genere (gender target)	6.87
	pesantezza (heaviness)	6.87
	stagione (season)	6.87
	taglio (cut)	6.87
	vestibilità ( )	6.87
	utilizzo (use)	6.14
	qualità (quality)	5.48
	destinatari (target group)	3.65
	luogo (location)	2.05
	composizione (composition)	2.74
	fascia di età (age group)	2.74
	produttore (producer)	2.34
forma (shape)	1.64	
anno (year)	1.64	
nome (name)	1.64	
categoria di appartenenza (category)	1.35	
nazionalità (nationality)	1.24	
dimensioni (dimensions)	0.66	

Table 75: Relevance Article of clothing (it)

Category	Attributes with $k \neq 0$	Relevance ( $k$ )
<i>Article of clothing</i>	gender intended for	24.04
	color	17.08
	material	16.42
	style	16.35
	fabric	13.74
	brand name	10.24
	price	8.66
	size	7.07
	manufacturer	7.01
	trademark	6.87
	name	6.56
	utility	5.48
	where available	5.48
	texture	4.67
	kind	4.10
	use	4.10
	type	3.77
	shape	3.28
	condition	2.74
	category	2.47
company	1.82	
year	1.35	
purpose/s	1.35	
weight	1.24	
description	1.04	
location	0.22	

Table 76: Relevance Article of clothing (eng)



<b>Category</b>	<b>Attributes with <math>k \neq 0</math></b>	<b>Relevance (<math>k</math>)</b>
<i>Object</i>	colore (color)	24.74
	materiale (material)	22.99
	forma (shape)	22.99
	funzione (function)	21.02
	peso	18.25
	nome (name)	14.78
	utilizzo (use)	12.29
	name of the object	9.85
	dimensioni (dimensions)	7.94
	funzionalità (functionality)	5.48
	odore (smell)	5.48
	epoca (epoch)	5.48
	misure (size)	4.10
	nazionalità (nationality)	2.47
	marca (brand)	2.34
	superficie (surface)	2.34
	produttore (producer)	2.34
	contenuto (content)	2.74
	spessore (thickness)	2.05
	luogo (location)	2.05
	posizione (position)	2.05
	caratteristiche (characteristics)	1.82
	autore (author)	1.64
	altezza (height)	1.49
proprietario (owner)	1.35	
prezzo (price)	0.95	
scopo (aim)	0.73	

Table 77: Relevance Object(it)

Category	Attributes with $k \neq 0$	Relevance ( $k$ )
<i>Object</i>	shape	16.42
	color	13.29
	id	11.68
	value	10.30
	name	9.85
	dimensions	9.34
	material	7.30
	features	6.87
	object type	6.87
	volume	6.87
	use	6.14
	purpose/s	6.77
	size	6.52
	conditions	5.48
	creator	5.48
	weight	4.95
	manufacturer	4.67
	texture	4.67
	creation date	4.10
	description	3.11
	owner/s	2.71
	category	2.47
	related to	2.74
	utility	2.74
	usage	2.34
	function	2.34
	price/s	1.24
	age	0.87
	location	0.65
	type	0.69

Table 78: Relevance Object (eng)

Category	Attributes with $k \neq 0$	Relevance ( $k$ )
<i>Tourist Location</i>	attrazioni (amenities)	21.93
	nome (name)	11.49
	possibilità di svago (amusements)	10.96
	numero di abitanti (number of inhabitants)	9.34
	posizione geografica (geographical position)	9.34
	luoghi vicini (near places)	7.01
	luogo (location)	6.14
	coordinate geografiche (lat, long) (geo coordinates)	5.47
	regione (region)	5.47
	paese (village)	4.67
	clima (climate)	4.67
	nazione (nation)	2.98
	prezzo (price)	2.85
	comune (municipality)	2.74
	numero di posti disponibili (number of accommodations)	2.74
	accoglienza (greeting)	2.74
	luminosità (luminance)	2.74
	musei (museums)	2.74
	temperatura (temperature)	2.74
	altitudine (altitude)	2.74
	attività (activity)	2.38
	estensione (extent)	2.34
	provincia (province)	2.34
	paesaggio (landse)	2.34
	caratteristiche ambientali (environmental characteristics)	2.05
	cap (zip code)	2.05
	accessibilità (access)	1.82
	collegamenti (connections)	1.82
	stato (state)	1.80
	descrizione (description)	1.74
	città (city)	1.20
	sito web (web site)	1.04
	scopo (aim)	0.73

Table 79: Relevance Tourist Location(it)

Category	Attributes with $k \neq 0$	Relevance ( $k$ )
<i>Tourist Location</i>	name	18.06
	attractions	13.74
	geo coordinates	11.50
	price/s	4.95
	area	4.67
	place	4.10
	post code	3.65
	gps coordinates	3.28
	distance from sea	2.74
	architectural style	2.74
	history	2.74
	neighboring cities	2.74
	places of interest	2.74
	time	2.74
	weather	2.74
	city	2.72
	category	2.47
	region	2.34
	post code	2.34
	population	2.34
	accessibility	2.34
	opening hours	2.34
	province	1.49
	country	1.41
	description	1.04
	state	0.95
	address	0.90
	location	0.86
	type	0.69
	web page url	0.60
size	0.54	

Table 80: Relevance Tourist Location (eng)

Category	Attributes with $k \neq 0$	Relevance ( $k$ )
<i>City</i>	numero abitanti (number of citizens)	56.05
	coordinate geografiche (geo coordinates)	23.72
	nome (name)	19.70
	regione (region)	18.25
	clima (climate)	14.01
	posizione geografica (geographical position)	11.68
	nazione (nation)	10.42
	monumenti (monuments)	10.30
	provincia (province)	9.34
	possibilità di svago (amusements)	8.22
	lingua (language)	8.22
	continente (continent)	8.22
	indirizzo (address)	7.21
	estensione (extension)	7.01
	storia (history)	6.14
	stato (state)	6.61
	industrie (industries)	6.87
	data di fondazione (date of foundation)	6.57
	arte (art)	6.87
	artigianato (crafts)	6.87
	capoluogo (capital)	6.87
	superficie (surface)	4.67
	paese (country)	4.67
	luoghi vicini)	4.67
	posizione (position)	4.10
	ambiente (enviroment)	2.74
	caratteristiche peculiari (distinguishing characteristics)	2.74
	conformazione (conformation)	2.74
	attrazioni (amenities)	2.74
	musei (museums)	2.74
	temperatura (temperature)	2.74
	dimensioni (dimensions)	2.65
	paesaggio (landscape)	2.34
	stile architettonico (architectural style)	2.34
	gestione (administration)	2.34
	cap (zip code)	2.05
	accessibilità (access)	1.82
	collegamenti (connections)	1.82
	luogo (location)	1.31
	città (city)	1.20
attività (activity)	0.79	
scopo (aim)	0.73	

Table 81: Relevance City (it)

Category	Attributes with $k \neq 0$	Relevance ( $k$ )
<i>City</i>	population	25.69
	name	21.34
	geo coordinates	11.50
	number of people	9.34
	language/s	9.12
	climate	6.87
	culture	6.87
	government	6.87
	area	4.67
	post code	3.65
	province	2.98
	state	2.85
	altitude	2.74
	architectural style	2.74
	history	2.74
	landmarks	2.74
	neighbouring city	2.74
	famous for	2.74
	place of interest	2.74
	planet	2.74
	time zone	2.74
	traffic	2.74
	weather	2.74
	region	2.34
	country	2.29
	location	2.15
	place	2.05
	gps coordinate	1.64
	size	1.63
	picture	1.49
	age	0.87

Table 82: Relevance City (eng)

Category	Attributes with $k \neq 0$	Relevance ( $k$ )
<i>Shop</i>	merce trattata (type of products sold)	48.08
	nome (name)	31.20
	orario (time)	30.15
	numero di dipendenti (number of employees)	11.50
	luogo (location)	8.77
	qualità dei prodotti (products quality)	6.87
	qualità del personale (staff quality)	6.87
	modalità di acquisto (purchase ways)	6.87
	giorno di chiusura (closing day)	5.48
	parcheggio (parking)	5.48
	via (street)	4.93
	dimensioni (dimensions)	3.97
	indirizzo (address)	3.60
	contatto/i (contact/s)	2.98
	servizio (service)	2.74
	accoglienza (greeting)	2.74
	clientela (clients)	2.74
	gestore (administrator)	2.74
	luminosità (luminance)	2.74
	modalità di pagamento (ways of payment)	2.74
	musica (music)	2.74
	sede (main office)	2.71
	proprietario (owner)	2.71
	qualità (quality)	2.34
	qualità del servizio (quality of service)	2.34
	settore (sector)	2.26
	servizi (services)	2.26
	numero di telefono (phone number)	2.26
	personale (staff)	2.05
	cap (zip code)	2.05
	coordinate geografiche (geo coordinates)	1.82
	accessibilità (access)	1.82
	livello (level)	1.82
	città (city)	1.80
	partita iva (VAT number )	1.49
	email	1.35
	colore (color)	1.24
	prezzo/i (price/s)	0.95
	descrizione (description)	0.87
	attività (activity)	0.79
scopo (aim)	0.73	
stato (state)	0.60	

Table 83: Relevance Shop (it)

Category	Attributes with $k \neq 0$	Relevance ( $k$ )
<i>Shop</i>	products sold	13.74
	name	13.13
	quality	5.48
	owner/s	5.42
	price/s	4.95
	beautiful	2.74
	chain	2.74
	landmarks	2.74
	opening hours	2.34
	address	2.09
	location	1.94
	color	1.90
	street	1.82
	web page url	1.80
	type	1.71
	geo coordinates	1.64
	number of employees	1.49
	service/s	1.49
	phone number	1.45
	size	1.09
city	0.54	
country	0.35	

Table 84: Relevance Shop (eng)



Category	Attributes with $k \neq 0$	Relevance ( $k$ )
<i>Hotel</i>	numero di stanze (number of rooms)	27.41
	numero di stelle (number of stars)	27.41
	nome (name)	24.63
	servizi (services)	12.45
	categoria di appartenenza (category)	10.84
	luogo (location)	7.89
	ristorante (restaurant)	6.87
	comfort	5.48
	indirizzo (address)	5.41
	prezzo (price)	4.75
	qualità (quality)	4.67
	luoghi vicini (near place)	4.67
	qualità del servizio (quality of service)	4.67
	fascia di prezzo (price range)	2.74
	clientela (clients)	2.74
	proprietario (owner)	2.71
	città (city)	2.40
	gestione (management)	2.34
	numero civico (civic number)	2.34
	stile architettonico (architectural style)	2.34
	numero di telefono (phone number)	2.26
	caratteristiche ambientali (environmental characteristics)	2.05
	personale (staff)	2.05
	dimensioni (dimensions)	1.98
	via (street)	1.64
	partita iva (VAT number)	1.49
	email	1.35
	colore (color)	1.24
	età (age)	1.24
	sito web (web site)	1.04
scopo (aim)	0.73	
stato (state)	0.60	

Table 85: Relevance Hotel (it)

Category	Attributes with $k \neq 0$	Relevance ( $k$ )
<i>Hotel</i>	name	32.84
	number of rooms	13.74
	rating	10.96
	amenities	10.30
	number of stars	10.30
	restaurant	10.30
	rooms	6.87
	city	3.81
	state	3.80
	price/s	3.71
	street	3.65
	post code	3.65
	address	3.58
	service/s	2.98
	type of cuisine	2.74
	chain	2.74
	facilities	2.74
	famous for	2.74
	fax number	2.74
	owner/s	2.71
	parking	2.34
	number of people	2.34
	location	1.94
	gps coordinates	1.64
	province	1.49
	number of employees	1.49
	phone number	1.45
	category	1.24
	country	1.23
	color	0.95
	email	0.95
age	0.87	
web page url	0.60	
size	0.54	

Table 86: Relevance Hotel (eng)

Category	Attributes with $k \neq 0$	Relevance ( $k$ )
<i>Restaurant</i>	nome (name)	21.34
	orario (time)	19.19
	specialità (specialty)	13.74
	piatti tipici (typical food)	10.30
	prezzo/i (price/s)	9.49
	numero di stanza (number of rooms)	8.22
	luogo (location)	7.45
	se attrezzato per diversamente abili (services for handicapped)	6.87
	accetta prenotazioni (reservation request)	6.87
	arredamento (furniture)	6.87
	vini (wines)	6.87
	categoria di appartenenza (category)	6.77
	giorno di chiusura (closing day)	5.48
	parcheggio (parking)	5.48
	fascia di prezzo (price range)	5.48
	numero di posti disponibili (number of accommodations)	5.48
	servizio (service)	5.48
	posizione (position)	4.10
	coordinate geografiche (geo coordinates)	3.65
	servizi (services)	3.39
	numero di dipendenti (number of employees)	3.28
	ambiente (environment)	2.74
	gestore (administrator)	2.74
	modalità di pagamento (ways of payment)	2.74
	musica (music)	2.74
	qualità del servizio (quality of service)	2.34
	gestione (management)	2.34
	qualità (quality)	2.34
	caratteristiche ambientali (environmental characteristics)	2.05
	accessibilità (access)	1.82
	livello (level)	1.82
	via (street)	1.64
	partita iva (VAT number)	1.49
dimensione/i (dimension/s)	1.32	
numero di telefono (phone number)	1.13	
scopo (aim)	0.73	
città (city)	0.60	
indirizzo (address)	0.60	

Table 87: Relevance Restaurant (it)

<b>Category</b>	<b>Attributes with <math>k \neq 0</math></b>	<b>Relevance (<math>k</math>)</b>
<i>Restaurant</i>	type of cuisine	52.08
	name	34.48
	chef	17.17
	kind of food	13.74
	price/s	13.60
	capacity	10.30
	rating	8.22
	gps coordinates	7.01
	hours of operation	6.87
	payment type	6.87
	performance/s	6.87
	quality of service	6.87
	reservation required	6.87
	address	5.97
	street	5.47
	service/s	4.46
	city	4.35
	owner/s	4.06
	number	3.28
	location	2.80
	time zone	2.74
	quality	2.74
	fax number	2.74
	opening hours	2.34
	parking	2.34
	accessibility	2.34
	email	1.90
	state	1.90
	post code	1.82
	type	1.71
	number	1.64
	picture	1.49
	province	1.49
	number of employees	1.49
	phone number	1.45
	category	1.24
	time	1.24
	web page url	1.20
	country	1.06
	description	1.04
	age	0.87
	size	0.54

Table 88: Relevance Restaurant (eng)

Category	Attributes with $k \neq 0$	Relevance ( $k$ )
<i>Place</i>	coordinate geografiche (geo coordinates)	29.19
	indirizzo (address)	16.82
	nome (name)	11.49
	altitudine (altitude)	10.96
	continente (continent)	8.22
	regione (region)	7.30
	tempo (weather)	6.87
	comune (municipality)	5.48
	posizione geografica (geographical position)	4.67
	clima (climate)	4.67
	provincia (province)	4.67
	nazione (nation)	4.46
	caratteristiche ambientali (environmental characteristics)	4.10
	colore (color)	3.71
	caratteristiche (characteristics)	3.65
	caratteristiche peculiari (peculiar characteristics)	2.74
	conformazione (conformation)	2.74
	comfort	2.74
	luogo (location)	2.63
	numero civico (civic number)	2.34
	numero abitanti (number of employees)	2.34
	estensione (extent)	2.34
	paese (village)	2.34
	paesaggio (landscape)	2.34
	accessibilità (access)	1.82
	collegamenti (connections)	1.82
	stato (state)	1.80
	via (street)	1.64
	età (age)	1.24
	servizi (services)	1.13
	attività (activity)	0.79
	scopo (aim)	0.73
dimensione/i (dimension/s)	0.66	
città (city)	0.60	

Table 89: Relevance Place (it)

Category	Attributes with $k \neq 0$	Relevance ( $k$ )
<i>Place</i>	geo coordinates	42.70
	name	21.34
	continent	13.74
	elevation	10.30
	distance from see	8.22
	street	7.30
	nearby place	6.87
	altitude	5.48
	post code	5.47
	city	5.44
	gps coordinates	4.93
	state	4.75
	region	4.67
	number	4.67
	population	4.67
	province	2.98
	color	2.85
	beautiful	2.74
	facilities	2.74
	planet	2.74
	traffic	2.74
	country	2.46
	parking	2.34
	number of people	2.34
	area	2.34
	accessibility	2.34
	service/s	1.49
	address	1.49
	description	1.04
	phone number	0.73
location	0.65	

Table 90: Relevance Place (eng)

### 4.3 A Rank-based Model of Relevance

The relevance measure proposed in the previous section does not consider the ranked nature of our data. We believe, however, that it is important to retain the ranked nature of participant responses. Participants were requested to list all attributes that they considered relevant to identify each category indicated. We assume that subjects have followed some order when they listed the attributes, starting with attributes more salient which occurred first to them and so on until attributes less important. Data models that ignore this crucial feature presume that later mentioned items (which some participants may have strained to come up with) represent the category as well as earlier ones. We retain that this is not the case. In many categories there were pairs of attributes that appeared on participants' lists with nearly equal frequency, though one typically preceded the other, indicating greater ease of recall or degree of prototypicality. In order to include this important aspect in our relevance measure, we propose a rank-based measure of relevance ( $\Omega$ ):

$$\Omega(c, f) = w(c, f) * \phi(c, f) * \psi(f)$$

where  $w(c, f)$  is a global factor ( $0 < w(c, f) \leq 1$ ) that combine the ranks of attribute  $f$  for the category  $c$  in the all the lists reported by participants.

$$w(c, f) = \frac{N}{\sum_{j=1}^{j=N} pos_j}$$

To give a numerical example, if the attribute  $f_1$  is listed by 5 subjects for the category  $c$  with the following ranks (pos ranges from  $j=1$  to  $j=5$ ): 1,3,2,1,1 we obtain:

$$w(c, f_1) = \frac{5}{1 + 3 + 2 + 1 + 1} = 0.62$$

Just to have a comparison, if another attribute  $f_2$  is listed by 5 subjects for the same category  $c$  but it is listed in the last positions of the lists (pos=5,4,6,4,5) the weight decreases:

$$w(c, f_2) = \frac{5}{5 + 4 + 6 + 4 + 5} = 0.20$$

$w(c, f) = 1$  when all subjects, that report the attribute  $f$  for the category  $c$ , list that attribute in first position.

$w(c, f) > 0$  for the simple reason that at least one subject reported the attribute considered, so the numerator (and consequently the denominator) of the formula can not be equal to 0.

#### 4.3.1 Relevance and Rank-based Relevance: a comparison

In order to compare the two measures we report the results on our data. We limit the presentation to the first seven attributes to make easier the comparison and render more salient possible chances in the order of attributes depending on relevance value. Moreover, to simplify the presentation of the results of the Italian version (it), we reported directly the translation of the attributes.

<b>Relevance (<math>k</math>)</b>		<b>Rank-based Relevance (<math>\Omega</math>)</b>	
party	35.63	name	15.4
political view	28.67	party	9.98
name	22.99	political view	8.89
age	21.02	age	6.94
surname	18.06	surname	6.86
education	14.78	role	4.32
role	14.88	profession	3.04

Table 91: Relevance measures: Politician (it)



<b>Relevance (<math>k</math>)</b>		<b>Rank-based Relevance (<math>\Omega</math>)</b>	
party	65.78	name	22.82
name	31.20	party	21.05
gender	16.42	gender	5.09
position/s held	14.60	surname	3.7
age	11.30	age	2.82
political view	10.30	position/s held	2.77
background	8.22	political view	2.58
surname	8.21	data elected	2.27

Table 92: Relevance measures: Politician (eng)

<b>Relevance (<math>k</math>)</b>		<b>Rank-based Relevance (<math>\Omega</math>)</b>	
company	21.93	name	21.35
name	21.34	surname	9.03
surname	18.06	company	6.58
experiences	9.12	ability	3.67
education	9.85	age	3.38
role/s	8.93	role	2.95
age	8.66	competence	2.64

Table 93: Relevance measures: Manager (it)

<b>Relevance (<math>k</math>)</b>		<b>Rank-based Relevance (<math>\Omega</math>)</b>	
name	26.27	name	13.66
experience/s	11.68	experience/s	4.2
reports	10.30	division/s	2.75
department/s	9.12	qualification/s	2.34
occupation	7.01	surname	2.12
budget	6.87	department/s	1.55
division/s	6.87	role	1.55
age	6.08	education	1.54

Table 94: Relevance measures: Manager (eng)

<b>Relevance (<math>k</math>)</b>		<b>Rank-based Relevance (<math>\Omega</math>)</b>	
teaching matter	54.94	teaching matter	18.13
name	21.34	name	12.6
institution where teaches	20.60	surname	6.57
publications	17.17	institution where teaches	4.74
surname	13.14	age	4.36
age	11.13	role	2.98
degree	8.22	education	2.63

Table 95: Relevance measures: Professor (it)

<b>Relevance (<math>k</math>)</b>		<b>Rank-based Relevance (<math>\Omega</math>)</b>	
university	37.77	name	28.97
name	34.48	university	8.31
publication	19.19	schools	5.48
research area	17.17	surname	4.11
department	14.60	department	3.5
education	9.48	research area	2.92
specialization	9.34	publication	2.49

Table 96: Relevance measures: Professor (eng)

<b>Relevance (<math>k</math>)</b>		<b>Rank-based Relevance (<math>\Omega</math>)</b>	
sport specialty	54.82	name	18.93
name	22.99	sport specialty	15.23
age	17.31	surname	5.51
surname	11.50	age	5.51
birth date	9.85	discipline	4.58
level	10.95	level	4.11
weight	7.30	birth date	3.11

Table 97: Relevance measures: Sportperson (it)

<b>Relevance (<math>k</math>)</b>		<b>Rank-based Relevance (<math>\Omega</math>)</b>	
type of sport	49.34	name	20.44
name	31.20	type of sport	15.86
team	17.17	team	4.29
birth date	11.50	gender	3.41
gender	14.78	awards	2.63
age	8.69	birth date	2.24
status professional	6.87	age	1.93

Table 98: Relevance measures: Sportperson (eng)

<b>Relevance (<math>k</math>)</b>		<b>Rank-based Relevance (<math>\Omega</math>)</b>	
films	34.34	name	21.74
name	24.63	surname	9.12
experiences	25.54	age	7.72
age	19.79	experiences	6.17
surname	16.42	films	5.92
nationality	13.60	education	5.75
birth date	11.50	birth date	3.83

Table 99: Relevance measures: Actor (it)

<b>Relevance (<math>k</math>)</b>		<b>Rank-based Relevance (<math>\Omega</math>)</b>	
name	26.27	name	18.39
films	19.19	films	4.03
birth date	11.50	birth date	4.03
gender	9.85	surname	2.96
awards	6.14	facial feature	2.74
surname	4.93	gender	2.56
role	4.14	character/s	2.34

Table 100: Relevance measures: Actor (eng)

<b>Relevance (<math>k</math>)</b>		<b>Rank-based Relevance (<math>\Omega</math>)</b>	
name	32.84	name	24.33
surname	27.92	surname	12.83
birth place	18.25	height	4.56
occupation	18.68	age	4.42
birth date	16.42	birth date	4.32
eyes color	13.70	weight	4.11
age	12.37	birth place	

Table 101: Relevance measures: Person (it)

<b>Relevance (<math>k</math>)</b>		<b>Rank-based Relevance (<math>\Omega</math>)</b>	
name	31.20	name	24.65
gender	22.99	gender	4.83
birth date	18.06	birth date	4.15
occupation	14.01	surname	3.74
religion	13.74	mother's name	2.96
marital status	10.96	partents name	2.78
mother's name	10.96	ethnicity	2.74

Table 102: Relevance measures: Person (eng)

<b>Relevance (<math>k</math>)</b>		<b>Rank-based Relevance (<math>\Omega</math>)</b>	
name	36.12	name	20.92
number of employees	14.78	structure	3.52
turnover	12.29	production	2.81
share capital	10.96	number of employees	2.77
output	10.30	people	2.74
structure	6.87	organization chart	2.73
products	6.87	sector	2.36

Table 103: Relevance measures: Company (it)

<b>Relevance (<math>k</math>)</b>		<b>Rank-based Relevance (<math>\Omega</math>)</b>	
name	24.63	name	23.09
ceo name	8.22	address	2.44
business type	8.19	business type	2.18
profits	7.01	profit/s	1.5
revenue	6.87	symbol	1.37
number of employees	5.95	history	1.37
logo	5.48	ceo name	1.32

Table 104: Relevance measures: Company (eng)

<b>Relevance (<math>k</math>)</b>		<b>Rank-based Relevance (<math>\Omega</math>)</b>	
name	27.91	name	20.63
members	27.47	members	7.85
objective/s	11.61	objective/s	4.89
number of members registered	7.01	associates	2.74
functions	6.87	field	2.74
sector	6.79	profession	2.34
main office	6.77	sector	2.04

Table 105: Relevance measures: Association (it)

<b>Relevance (<math>k</math>)</b>		<b>Rank-based Relevance (<math>\Omega</math>)</b>	
name	21.34	name	16.23
objective/s	16.42	objective/s	4.27
members	11.68	leader/s	4.11
activity	11.68	members	3.62
date of foundation	9.12	domain	3.43
leader/s	8.22	international	2.75
international	6.87	date of foundation	2.37

Table 106: Relevance measures: Association (eng)

<b>Relevance (<math>k</math>)</b>		<b>Rank-based Relevance (<math>\Omega</math>)</b>	
faculties	30.91	name	20.12
name	22.99	faculties	11.13
number of students	17.17	number of students	4.09
courses	16.35	courses	3.18
professors	13.74	place	2.9
degree course	10.30	department/s	2.24
department/s	8.22	structure	2.06

Table 107: Relevance measures: University (it)

<b>Relevance (<math>k</math>)</b>		<b>Rank-based Relevance (<math>\Omega</math>)</b>	
name	26.27	name	24.07
number of students	19.19	faculty/ies	3.76
faculty/ies	16.35	courses	3.56
courses	13.70	number of majors	3.43
department/s	9.12	number of students	3.26
international programs	6.87	department	2.01
number of majors	6.87	city	1.79

Table 108: Relevance measures: University (eng)

<b>Relevance (<math>k</math>)</b>		<b>Rank-based Relevance (<math>\Omega</math>)</b>	
political view	16.38	nation	10.42
duration	13.74	party	6.26
party/s	10.96	political view	5.96
ministries	10.30	duration	5.49
ministers	10.30	people	2.74
premier	10.30	premier	2.58
nation	10.42	ministries	2.21

Table 109: Relevance measures: Government (it)

<b>Relevance (<math>k</math>)</b>		<b>Rank-based Relevance (<math>\Omega</math>)</b>	
name	11.49	party	8.22
head	9.34	name	4.71
members	9.34	head	3.73
party	8.22	level	2.75
leaders	8.22	members	2.24
level	6.87	leaders	2.06
characteristics	5.48	country	1.87

Table 110: Relevance measures: Government (eng)

<b>Relevance (<math>k</math>)</b>		<b>Rank-based Relevance (<math>\Omega</math>)</b>	
name	26.27	name	17.52
number of employees	8.21	field	2.74
clients	6.87	clients	2.29
sector	4.53	objective/s	1.74
objective/s	4.36	legal form	1.37
staff	4.10	areaof work	1.37
address	3.60	address	1.14

Table 111: Relevance measures: Agency (it)

<b>Relevance (<math>k</math>)</b>		<b>Rank-based Relevance (<math>\Omega</math>)</b>	
name	21.34	name	12.6
number of employees	7.44	specialization	2.34
president	6.87	specialization	2.34
specialization	4.67	product/s	1.87
profit/s	4.67	number of employees	1.49
product/s	4.67	acronym	1.37
funding	4.67	short name	1.37

Table 112: Relevance measures: Agency (eng)

<b>Relevance (<math>k</math>)</b>		<b>Rank-based Relevance (<math>\Omega</math>)</b>	
name	27.91	name	14.83
objective/s	12.34	objective/s	4.66
members	11.68	sector	3.88
sector	9.05	members	3.24
date of foundation	8.21	legal form	2.74
main office	8.13	date of foundation	2.28
mission	7.01	nation	2.23

Table 113: Relevance measures: Organization (it)

<b>Relevance (<math>k</math>)</b>		<b>Rank-based Relevance (<math>\Omega</math>)</b>	
name	21.34	name	19.85
business type	6.14	character	4.67
objective/s	4.93	business type	2.33
character/s	4.67	role	2.07
head	4.67	scale	1.17
product/s	4.67	product/s	1.03
business registration number	4.10	sector	1.02

Table 114: Relevance measures: Organization (eng)

<b>Relevance (<math>k</math>)</b>		<b>Rank-based Relevance (<math>\Omega</math>)</b>	
topic	25.54	topic	11.18
speakers	16.45	title	5.87
participants	16.42	participants	4.56
date	11.13	speakers	3.8
needs	10.30	date	3.45
year	10.30	location	2.63
duration	9.85	time	2.38

Table 115: Relevance measures: Conference (it)

<b>Relevance (<math>k</math>)</b>		<b>Rank-based Relevance (<math>\Omega</math>)</b>	
name	22.32	name	16.74
organizers	12.77	date	3.46
date	12.37	year	3.39
chair/s	10.30	title	3.27
sessions	10.30	organizers	2.94
topic/s	10.24	topics	2.87
sponsor/s	8.19	purpose	2.71

Table 116: Relevance measures: Conference (eng)

<b>Relevance (<math>k</math>)</b>		<b>Rank-based Relevance (<math>\Omega</math>)</b>	
topic	20.07	topic	7.12
time	17.86	time	6.7
participants	16.42	participants	5.66
date	11.13	date	4.77
location	7.01	sector	3.39
speaker	5.48	location	3.03
recipients	3.65	day	2.05

Table 117: Relevance measures: Meeting (it)

<b>Relevance (<math>k</math>)</b>		<b>Rank-based Relevance (<math>\Omega</math>)</b>	
time	29.76	subject	14.78
date	19.79	time	6.68
topic/s	18.43	date	6.13
participants	16.08	name	4.61
agenda	13.74	topic/s	4.61
minutes	10.30	participants	2.89
subject	9.85	schedule	2.74

Table 118: Relevance measures: Meeting (eng)

<b>Relevance (<math>k</math>)</b>		<b>Rank-based Relevance (<math>\Omega</math>)</b>	
topic	21.89	topic	11.42
duration	11.50	exhibitors	6.18
artists	10.96	title	4.55
exhibitors	10.30	organizers	3.71
title	9.85	date	3.19
artist	9.34	name	3.65
date	8.66	artist	3.13

Table 119: Relevance measures: Exhibition (it)

<b>Relevance (<math>k</math>)</b>		<b>Rank-based Relevance (<math>\Omega</math>)</b>	
name	11.90	name	6.31
time	8.93	subject	3.28
date	7.42	time	3.19
start date	7.01	start date	1.75
end date	4.93	content	1.37
subject	3.28	role	1.04
duration	2.98	end date	0.99

Table 120: Relevance measures: Exhibition (eng)

<b>Relevance (<math>k</math>)</b>		<b>Rank-based Relevance (<math>\Omega</math>)</b>	
date	19.79	title	7
actors	18.68	date	6.6
duration	18.06	duration	6.21
time	16.37	actors	4.98
title	14.78	time	4.62
location	11.83	name	3.65
direction	10.30	location	3.63

Table 121: Relevance measures: Show (it)

<b>Relevance (<math>k</math>)</b>		<b>Rank-based Relevance (<math>\Omega</math>)</b>	
actors	17.17	name	5.49
name	13.39	actors	4.98
producer/s	10.30	protagonist	3.43
time	7.44	performer/s	2.75
director/s	7.01	length	2.74
performer/s	6.87	producer/s	2.37

Table 122: Relevance measures: Show (eng)



<b>Relevance (<math>k</math>)</b>		<b>Rank-based Relevance (<math>\Omega</math>)</b>	
people involved	25.69	people involved	7.25
entities involved	24.04	causes	6.87
damages	13.74	entities involved	5.43
dynamics	13.74	location	4.58
vehicles involved	10.30	dynamics	3.66
time	11.90	time	3.28
date	8.66	damages	3.23

Table 123: Relevance measures: Incident (it)

<b>Relevance (<math>k</math>)</b>		<b>Rank-based Relevance (<math>\Omega</math>)</b>	
time	20.83	time	7.79
date	11.13	date	5.23
people involved	10.96	people involved	3.18
participants	7.42	participants	2.37
causes	7.01	start time	1.54
severity	6.87	type of incident	1.51
type of incident	6.87	year	1.35

Table 124: Relevance measures: Incident (eng)

<b>Relevance (<math>k</math>)</b>		<b>Rank-based Relevance (<math>\Omega</math>)</b>	
date	18.55	date	8.43
participants	13.14	name	4.47
name	12.77	participants	2.44
duration	8.21	location	2.35
time	7.44	time	2.19
motivations	6.87	topic	1.82
attendants	6.87	attendants	1.72

Table 125: Relevance measures: Event (it)

<b>Relevance (<math>k</math>)</b>		<b>Rank-based Relevance (<math>\Omega</math>)</b>	
time	17.86	date	11.08
date	17.31	participants	9.89
name	10.41	time	7.12
participants	9.89	name	5.62
repetition	6.87	what	1.87
purpose/s	6.77	start time	1.64
outcome	5.48	location	1.63

Table 126: Relevance measures: Event (eng)

<b>Relevance (<math>k</math>)</b>		<b>Rank-based Relevance (<math>\Omega</math>)</b>	
type of sport	27.41	type of sport	15.23
name	14.59	name	10.62
date	13.60	start date	5.48
location	7.01	date	4.99
time	5.95	location	2.55
start date	5.48	nation	1.49
participation ways	5.48	participants	1.38

Table 127: Relevance measures: Sports Event (it)

<b>Relevance (<math>k</math>)</b>		<b>Rank-based Relevance (<math>\Omega</math>)</b>	
type of sport	30.15	type of sports	22.01
stadium	10.30	date	3.46
date	12.37	players	3.43
time	7.44	name	2.62
winners	6.87	game	2.75
game	6.87	stadium	2.58
players	6.87	winners	2.27

Table 128: Relevance measures: Sports Event (eng)

<b>Relevance (<math>k</math>)</b>		<b>Rank-based Relevance (<math>\Omega</math>)</b>	
use	22.52	name	9.46
name	18.06	use	6.7
price	14.24	price	3.81
color	11.13	components	3.43
brand	9.34	shape	2.96
features	9.12	brand	2.87
weight	9.12	description	2.61

Table 129: Relevance measures: Product (it)

<b>Relevance (<math>k</math>)</b>		<b>Rank-based Relevance (<math>\Omega</math>)</b>	
manufacturer	21.02	name	9.46
price/s	14.84	use	9.17
name	14.78	manufacturer	8.02
use	14.33	brand name	4.61
warranty	10.30	price	4.01
availability	6.87	id	2.34
brand name	6.14	warranty	2.06

Table 130: Relevance measures: Product (eng)

<b>Relevance (<math>k</math>)</b>		<b>Rank-based Relevance (<math>\Omega</math>)</b>	
author	37.77	author	17.73
location	24.57	artist	9.34
style	18.68	period	7.73
creation date	17.17	name	5.26
technique	17.17	location	5.17
title	14.78	creation date	5.05
material	14.78	style	4.98

Table 131: Relevance measures: Artwork (it)

<b>Relevance (<math>k</math>)</b>		<b>Rank-based Relevance (<math>\Omega</math>)</b>	
artist/s	30.91	artists	21.33
creation date	18.43	author/s	5.25
style	14.01	name	4.93
material	10.95	creation date	3.69
author/s	7.01	title	2.9
display location	6.87	style	2.66
technique	6.87	material	2.41

Table 132: Relevance measures: Artwork (eng)

<b>Relevance (<math>k</math>)</b>		<b>Rank-based Relevance (<math>\Omega</math>)</b>	
number of floors	37.77	location	15.42
location	32.76	number of floors	10.65
mq	20.60	height	7.5
height	16.37	mq	6.51
date of creation	13.74	position	4.1
color	12.37	color	2.81
use	12.29	dimension	2.54

Table 133: Relevance measures: Building (it)

<b>Relevance (<math>k</math>)</b>		<b>Rank-based Relevance (<math>\Omega</math>)</b>	
architect	20.60	height	4.17
number of floors	17.17	number of floors	4.12
height	14.90	name	.28
name	13.13	architect	2.88
architectural style	10.30	architectural style	1.34
floors	10.30	gps coordinates	2.46
use	8.19	address	2.42

Table 134: Relevance measures: Building (eng)

<b>Relevance (<math>k</math>)</b>		<b>Rank-based Relevance (<math>\Omega</math>)</b>	
editor	61.81	editor	23.67
number of pages	54.94	title	23.38
author	44.34	author	21.38
title	36.13	number of pages	11.13
publication date	30.91	publication date	5.68
ISBN	24.04	language/s	5.48
edition	13.74	ISBN	5.26

Table 135: Relevance measures: Book (it)

<b>Relevance (<math>k</math>)</b>		<b>Rank-based Relevance (<math>\Omega</math>)</b>	
author/s	46.71	author/s	22.78
publisher	44.64	publisher	10.55
ISBN	35.63	ISBN	10.07
year of publication	27.47	title	8.45
number of pages	24.04	number of pages	7.32
title	13.79	year of publication	6.46
edition	10.30	name	4.93

Table 136: Relevance measures: Book (eng)

<b>Relevance (<math>k</math>)</b>		<b>Rank-based Relevance (<math>\Omega</math>)</b>	
sizes	72.11	sizes	27.53
brand	35.03	color	12.7
color	33.39	brand	9.91
fabric	27.47	fabric	8.45
model	17.17	model	4.77
price	11.39	style	3.4
style	9.34	season	2.75

Table 137: Relevance measures: Article of clothing(it)

<b>Relevance (<math>k</math>)</b>		<b>Rank-based Relevance (<math>\Omega</math>)</b>	
color	17.08	trademark	6.87
material	16.42	gender intended for	6.47
style	16.35	color	6.03
fabric	13.74	utility	5.48
brand name	10.24	name	4.38
price	8.66	style	4.09
size	7.07	material	3.69

Table 138: Relevance measures: Article of clothing(eng)

<b>Relevance (<math>k</math>)</b>		<b>Rank-based Relevance (<math>\Omega</math>)</b>	
color	24.74	name	9.5
material	22.99	shape	8.7
shape	22.99	material	7.32
function	21.02	color	6.97
weight	18.25	function	5.73
name	14.78	weight	4.56
use	12.29	use	2.63

Table 139: Relevance measures: Object (it)

<b>Relevance (<math>k</math>)</b>		<b>Rank-based Relevance (<math>\Omega</math>)</b>	
shape	16.42	shape	5.47
color	13.29	id	4.49
id	11.68	color	4.04
value	10.30	name	3.7
name	9.85	creator	3.65
dimensions	9.34	object type	3.43
material	7.30	material	3.24

Table 140: Relevance measures: Object (eng)

<b>Relevance (<math>k</math>)</b>		<b>Rank-based Relevance (<math>\Omega</math>)</b>	
amenities	21.93	name	10.06
name	11.49	geographical position	7.47
fun opportunities	10.96	amenities	4.39
number of inhabitants	9.34	location	3.58
geographical position	9.34	number of inhabitants	2.87
near places	7.01	fun opportunities	2.74
location	6.14	activities	2.38

Table 141: Relevance measures: Tourist Location (it)

<b>Relevance (<math>k</math>)</b>		<b>Rank-based Relevance (<math>\Omega</math>)</b>	
name	18.06	name	7.64
attractions	13.74	geo coordinates	3.09
geo coordinates	11.50	place of interest	2.74
price/s	4.95	place	2.73
area	4.67	attractions	2.39
place	4.10	gps coordinates	2.19
post code	3.65	post code	1.82

Table 142: Relevance measures: Tourist Location (eng)

<b>Relevance (<math>k</math>)</b>		<b>Rank-based Relevance (<math>\Omega</math>)</b>	
number of citizens	56.05	number of citizens	14.16
geo coordinates	23.72	name	11.82
name	19.70	geographical position	8.34
region	18.25	language	8.22
climate	14.01	geo coordinates	7.71
geographical position	11.68	region	5.53
nation	10.42	nation	4.29

Table 143: Relevance measures: City (it)

<b>Relevance (<math>k</math>)</b>		<b>Rank-based Relevance (<math>\Omega</math>)</b>	
population	25.69	population	7.44
name	21.34	name	6.31
geo coordinates	11.50	geo coordinates	3.22
number of people	9.34	number of people	3.11
language/s	9.12	post code	2.43
climate	6.87	place	2.05
culture	6.87	culture	1.72

Table 144: Relevance measures: City (eng)

<b>Relevance (<math>k</math>)</b>		<b>Rank-based Relevance (<math>\Omega</math>)</b>	
type of products sold	48.08	type of products sold	17.71
name	31.20	name	17.44
tim	30.15	time	6.03
number of employees	11.50	location	3.81
location	8.77	staff quality	2.75
products quality	6.87	service	2.74
staff quality	6.87	products quality	2.29

Table 145: Relevance measures: Shop (it)

<b>Relevance (<math>k</math>)</b>		<b>Rank-based Relevance (<math>\Omega</math>)</b>	
products sold	13.74	name	9.55
name	13.13	products sold	4.99
quality	5.48	quality	2.19
owner/s	5.42	price/s	1.98
price/s	4.95	geo coordinates	1.64
beautiful	2.74	landmarks	1.37
chain	2.74	beautiful	1.37

Table 146: Relevance measures: Shop (eng)

<b>Relevance (<math>k</math>)</b>		<b>Rank-based Relevance (<math>\Omega</math>)</b>	
number of rooms	27.41	name	16.07
number of stars	27.41	number of stars	7.61
name	24.63	number of rooms	6.53
services	12.45	location	4.73
category	10.84	category	4.33
location	7.89	address	4.06
restaurant	6.87	services	2.85

Table 147: Relevance measures: Hotel (it)

<b>Relevance (<math>k</math>)</b>		<b>Rank-based Relevance (<math>\Omega</math>)</b>	
name	32.84	name	23.46
number of rooms	13.74	restaurant	5.15
rating	10.96	room	2.29
amenities	10.30	number of rooms	2.11
number of stars	10.30	gps coordinates	1.64
restaurant	10.30	number of stars	1.4
rooms	6.87	street	1.22

Table 148: Relevance measures: Hotel (eng)

<b>Relevance (<math>k</math>)</b>		<b>Rank-based Relevance (<math>\Omega</math>)</b>	
name	21.34	name	18.5
time	19.19	specialty	3.92
specialty	13.74	location	3.17
typical food	10.30	time	2.98
price/s	9.49	price	2.88
number of rooms	8.22	furniture	2.29
location	7.45	typical food	2.06

Table 149: Relevance measures: Restaurant (it)

<b>Relevance (<math>k</math>)</b>		<b>Rank-based Relevance (<math>\Omega</math>)</b>	
type of cuisine	52.08	name	19.57
name	34.48	type of cuisine	12.37
chef	17.17	kind of food	4.99
kind of food	13.74	address	2.65
price/s	13.60	quality of service	1.98
capacity	10.30	chef	1.87
rating	8.22	gps coordinates	1.64

Table 150: Relevance measures: Restaurant (eng)

<b>Relevance (<math>k</math>)</b>		<b>Rank-based Relevance (<math>\Omega</math>)</b>	
geo coordinates	29.19	geo coordinates	10.15
address	16.82	name	8.05
name	11.49	address	3.36
altitude	10.96	geographical position	3.11
continent	8.22	continent	3.08
region	7.30	conformation	2.74
weather	6.87	altitude	2.58

Table 151: Relevance measures: Place (it)

<b>Relevance (<math>k</math>)</b>		<b>Rank-based Relevance (<math>\Omega</math>)</b>	
geo coordinates	42.70	geo coordinates	15.21
name	21.34	name	11.1
continent	13.74	gps coordinates	3.7
elevation	10.30	elevation	3.43
distance from see	8.22	post code	3.28
street	7.30	altitude	2.74
nearby place	6.87	continent	2.29

Table 152: Relevance measures: Place (eng)



## 4.4 Final considerations

In this document we presented a general procedure to estimate the relevance of an attribute for the entity identification. Our goal was not producing final results, clearly limited from the initial ontological choices and from the size of the sample tested, but giving a methodological contribution to research areas that adopt attributes as their representational currency. In particular in the context of entity-based systems this procedure could provide useful insights for the development of entity matching algorithm. The idea could be exploiting the relevance values as weights to assign to attributes. Another important advantage of the procedure proposed pertains to its empirically grounded nature. On the one hand this allows to conduct investigations in more specialized sectors, analyzing types of entity relevant for that specific contexts. On the other hand it could be generalized gathering the attribute frequency directly from the users which use the entity-based systems (for example when they add new entities or when they search for entities). In this revised version, the procedure could also be capitalized by life cycle management applications, exploiting the possibility of following if and how the relevance of some attributes could be change in time. Finally the results collected from users from different cultures or languages can suggest different strategies when differences emerge patently.

## References

- [1] A. Gangemi, N. Guarino, C. Masolo, A. Oltramari, and L. Schneider. Sweetening ontologies with dolce. In *Lecture Notes In Computer Science, Proceedings of the 13th International Conference on Knowledge Engineering and Knowledge Management. Ontologies and the Semantic Web*, volume 2473, 2002.
- [2] C. Masolo, S. Borgo, A. Gangemi, N. Guarino, and Al. Oltramari. *WonderWeb Deliverable D18 Ontology Library (final)*. 2003.
- [3] C. Matuszek, J. Cabral, M. Witbrock, and J. DeOliveira. An introduction to the syntax and content of cyc. In *Proceedings of the 2006 AAAI Spring Symposium on Formalizing and Compiling Background Knowledge and Its Applications to Knowledge Representation and Question Answering.*, Stanford, March 2006.
- [4] K. McRae, G.S. Cree, M.S. Seidenberg, and C. McNorgan. Semantic feature production norms for a large set of living and nonliving things. *Behavior Research Methods, Instrument and Computers*, 37:547–559, 2005.
- [5] I. Niles and A. Pease. Towards a standard upper ontology. In Chris Welty and Barry Smith, editors, *Proceedings of the 2nd International Conference on Formal Ontology in Information Systems*, Ogunquit, Maine, October 17-19, 2001.
- [6] A. Oltramari, A. Gangemi, N. Guarino, and C. Masolo. Restructuring wordnet’s top-level: The ontoclean approach. In *n K. Simov (ed.) Workshop Proceedings of OntoLex’2, Ontologies and Lexical Knowledge Bases*, Las Palmas, Spain, May 27, 2002.
- [7] E. Rosch, C.B. Mervis, W. Gray, D.Johnson, and Boyes-P. Braem. Basic objects in natural categories. *Cognitive Psychology*, 8:382–439, 1976.
- [8] G. Sartori and L. Lombardi. Semantic relevance and semantic disorders. *Journal of Cognitive Neuroscience*, 16:439–452, 2004.