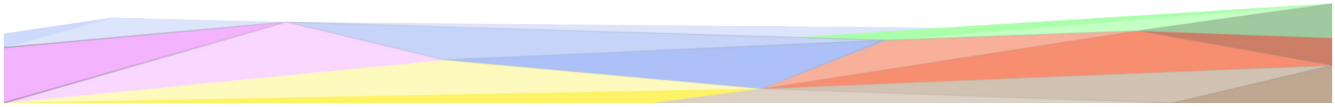
Dipartimento di Ingegneria e Scienza dell'Informazione

– KnowDive Group –

Tourist Events

Document Data:	Reference Persons:
October 15, 2020	Erdenebileg Byambadorj
	Eva Agrawal



Contents

Revision History:

Revision	Date	Author	Description of Changes

1 Knowledge Graph Codebook

The first of the two sections, in the current document, contains the codebook of the whole KG (Knowledge Graph), including the description of all the data and information that it contains.

1.1 Knowledge Graph general description

This sub section aims to give a general description of the KG, reporting:

- the context/domain in which the KG lives and works;
- *The Problem* the KG aims to solve;
- How the KG can solve *The Problem*

1.2 Data level

The data level section aims to describe in details the (final version of) datasets collected and managed by the KG, with a description of each variable involved.

1.2.1 Datasets general details

In this section are reported the metadata at datasets level, so the metadata regarding the sources, the authors, the collection methods, and so on.

1.2.2 Datasets metadata documentation

In this section are reported the metadata at dataset attribute level, through a description of each variable involved in the datasets collected, specifying the variable types, meanings, value-set (possible values), and every other meaningful variable information.

1.3 Ontology level

The ontology level section aims to describe the underlying KG ontology, through the description of its elements at each level, reporting so the language, conceptual and schema resources used within it.

1.3.1 Ontology general details

This first sub section of the ontology level description, report the general details such as authors, sources and the description of external ontology eventually adopted to generate the final one.

1.3.2 Ontology metadata documentation

In this section instead, are reported the more specific metadata describing the single elements of the ontology (terms, concepts, ETypes and relations).

1.4 Knowledge Graph Evaluation

In the final section of this first chapter, the KG Evaluation is reported. It aims to describe, through specific metrics, the quality of the overall KG on different aspect, like domain coverage, usability, domain representation, and other meaningful aspects.

2 Knowledge Graph Development Process

The second chapter of this document aims to describe, in a detailed way, the KG development process. The sections below describe each phase of the KG building project, reporting for each phase, the description of the datasets and their evolution respect the previous phases, the schema construction which will generate the KG ontology in the end, as well as the description of the procedures adopted to manage the data and finally achieve those results. Moreover for each phase is reported an evaluation section, which aims to evaluate the quality of the results achieved at the end of each phase.

2.1 Scope Definition

2.1.1 Purpose

Aside from experiencing beautiful sights of nature, a tourist is supposed to be amused and entertained by local festivals, concerts, visiting historical areas, monuments, restaurants and any other engaging events organized inside the territory he or she is travelling. To pick up the best set of events for a tourist to participate can be achieved by inspecting the adequate reviews from the previous tourists for each event. Nonetheless, this gives rise to two significant challenges. First, the tourist, for a given period, has to stack up the activities corresponding to his or her interest, as a result of choosing the most compatible ones from each of the event categories, such as points of interest to visit, films to watch, restaurants to dine at, community meetings, and other events. Second, assuming the combination the tourist favors the most is found, optimal arrangement of those events in the best possible chronological order is needed. For instance, one cannot be expected to leave for another event while the current activity is still ongoing or the location of the next one is too far. With a simple observation, it is effortless to tell a pair of combinatorial and permutation tasks by a machine could solve these manual optimizations as long as there is a coherently schematized knowledge graph associated to a different set of datasets concerning diverse event contexts. Therefore, this project is aimed at coming up with a solution to the data heterogeneity problem described above by means of designing a knowledge graph which will integrate data about various tourism events, basing on the locality, time, individual behaviour patterns and interaction preferences in order for the tourist to easily enlist a perfect sequence of events.

2.1.2 Scope of work

Considering the amount of data we have found, size of the end application, time we are assigned, we have decided to integrate datasets, on the scope of only the city of Trento, composed of points of interests, restaurants, shops, hotels, monuments to visit, their ratings and reviews from TripAdvisor, movies from cinemas, special public events, festivals, concerts and sport activities.

2.1.3 Scenarios of usages

Tourist events are one of the exciting experiences everyone would love to explore during their lifetime. Getting to feed your eyes with what you may have never seen or what your eyes love to see most, one can receive a sense of refreshment and containment at a personal level depending on an individual idiosyncrasies. However, this experience can be more memorable if the answers of what, when, why and how to attain them are fully tackled. Let us discover more with the help of the following personas we present:

- **Sofia:** An international student to study at University of Trento through the Erasmus+ program, being always busy on weekdays throughout the day. She always decides to go the theater near the train station of Trento, to watch night films on Friday nights as a way to refreshing her mind after a long week of class work. She is fond of two activities at weekend: patronizing flea markets to collect merchandises, which would later become reminiscent of the particular local areas she had visited to and being entertained by different theatrical performances depending on the mood she is in. For the five months period, the catalogues of merchandises up for sale, whose arrangements are organized on a certain day in a specific commune in Trentino, and the typologies of theatrical events with the associated descriptions would facilitate her scheduling while keeping her academic development intact. In addition, her lack of familiarity with the language and the country's historical contexts would deter her comprehensive understanding of some performances, a few categories such as ballet, opera or circus events which don't require prior information and are mostly universally interpreted activities could be optimal. Lastly, while attending events wherein varieties of local hand-made products not to be found on websites like amazon are available for sale, she is also captivated by cuisine exhibitions being organized close to the area where she is.

- **Tojo:** a Japanese novelist, a tourist who happens to come to the province of Trentino Alto-Adige in hopes of discovering fresh sources of materials to serve as an essential foundation in his creation of new character personalities and subtle environmental atmosphere. Unfortunately, he is stay in here all November, leaving him no choice, but to detect art exhibitions to be organized in this region during his stay, in diversity of forms like monuments, scientific installations, abstract arts, and geometric abstractions, from different times such as S.P.Q.R legacies, religious extremism of dark ages, renaissance, enlightenment and contemporary arts. On top of that, he is expecting to participate in guided excursions like a bicycle trip or a hiking around the mountainous areas in an effort not to lose the precious time exploring which could be the most favourable paths, and to receive interactive explanations from the tour guide regarding the places of which he would stop to take pictures. Although the ideal travel routes he prioritizes would be better off being compact, the randomness of context would play a huge part in giving him an authentic novel inspirations

- **Davide:** Davide is a single working man he lives in Archo at the heart of Garda Trentino, being a single man he boasts his social life away from work by attending most of the events organised within as a way of beating boredom and making friends. Staying in Archo has even made this easier and more interesting since it is one the destinations for one of the refreshing holiday events. However, this leaves him with a dilemma of which event to attend as regards time, venue ,ambience hence he is always confused and sometimes makes wrong decisions. He prefers going for mountain biking over the Saturday since it is his day off from work. One day he missed it yet it's one of his favourite events, just because he forgot the start time. Being a curious passion Davide wants to know more about his culture and history, so once in while to attends cultural exhibitions where he is able to have a taste of different people's way of life. Christmas shopping is one of the things he enjoys doing and since he prefers celebrating this season with his family which is far away from the prominent

Christmas markets, he decides to do his shopping earlier and takes the items shopped along with him. While in the eve of Christmas Davide enjoys taking having a memorable time with his family and friends so he organises them and takes them out on family festivals within his locality.With this he has to find out the most convenient and available means of transport for all of them to and from the event.

- **Antonio:** A married man with three kids, he is an Italian native working and living in Trento together with his family, his wife works in Rovereto. Every after three months the couple gets work leaves to spend some precious time together with their children, after a long period of hectic schedules. They prefer attending most of the tourist events available during that period when they are free as a way of bonding more with their children and creating precious memories to remember. His wife shopping from flea markets mostly clothing for their children and some domestic stuff for the family which she does once every two weeks. During the Christmas period they enjoy going to shop in different markets in hunt for the best and affordable items to make their celebrations colourful and fun. The family enjoys movie nights too most especially over on Saturdays, however they prefer a movie which does not take too long to become boring and besides that they don't wish to travel with their children late at night during the cold. Sometimes they invite their friend their international friends to join them, so for this case they would prefer to watch a movie that is translated to English or fully in English since they too understand English. As a way of bonding with other families they once in while attend family festivals were their kids can get to interact with others and they also share family experiences and learn from each other.

Name	Age	Interests	Usages	Description
Sofia	20	Movies, Market exhibitions, dance and opera performances	Detecting public expos displayed nearby, categorizing theatrical performances set by different aspects	Sophia, as expected from an international student, cannot speak Italian that well, also she's devoid of any cultural background. So, she could use the system for determining which performances are more universally understandable as well as quite enjoyable within the tight budget of a regular student.
Tojo	42	Guided tour, museum and other art exhibitions, gallery tour	Using the system to identify most productive, yet the least time-consuming guided tours	Optimization of his time can be easily resolved by the system. But he also might want to use the typology of event and a certain keyword to suit the application for his best need.

Davide	30	Italian cultural events, family meetings, public events	Event division by topics of discussion, detecting family meetings, overcoming language barrier	To be enjoyed by public events doesn't require anything linguistic. However, it is substantial that people coming to the meetings whose topics might interest him can comprehend and communicate to him.
Antonio	39	Weekly, monthly or flea market expo, hand-made product demonstration, movies	Searching local expos, cataloguing movie list for his family	The system may be utilized for finding variety of types of movies for all age groups within a given period. Also, it would be useful to pick up shows and entertainment for children while on a tour around the city.

2.2 Inception

This section is dedicated to the Inception phase description. Here are reported the initial definitions for CQs (Competency Queries), initial datasets collected and the relative metadata. For each of those elements the procedures and the tools adopted to achieve the results, have to be reported in the sections below.

2.2.1 Competency Queries

In order to round up a collection of **etypes** and their properties, we will make use of the following table of competency queries as to what could possibly be question instances parameterized in terms of generic questions by persona and its systemized implementations.

Persona	**№**	**Generalized Question**	**Action**
Sofia	1.1	Give the list of all flea market events being organized in Trentino next day.	Retrieve a set of all records categorized under flea market given the current date and no constraint for the selection of commune.
Sofia	1.2	Process the trade fair market available nearby	Return the trade fair market to be available within the closest vicinity of Trento geographically by tomorrow
Sofia	1.3	Process the list of weekly market being organized at the weekends	Some markets operate on one specific day of weeks. Restriction on a day would definitely reduces the number of choices to visit.
Sofia	1.4	Process the markets to be in attendance until the evening.	Return the list of markets in descending order of being closed the latest to the earliest before 8:00PM.

Sofia	1.5	Process whether there are ballet performances in Auditorium Fausto Melotti Rovereto	Suggest, if there exists any, entertainment events with type equal to the number corresponding to 'ballet' and the property nameTheatre set to 'Auditorium Fausto Melotti Rovereto'.
Sofia	1.6	Process and order the stack of theatrical performances in the next week according to their durations in the descending order	After the selection of entertainment events within the interval of next Monday and next Sunday, there would be subtraction of startDate from endDate to extract durations, which would be ordered and returned.
Sofia	1.7	Process the average costs associated with generic event organizers.	Group all the events according to their organizers in order to aggregate and average their costs for every organizer.
Sofia	1.8	Process the top 5 exhibitions scheduled in terms of spatial and temporal proximity	Extract and retrieve the top 5 records containing the pairs of differences of the current time and location from the ones of the events whose category falls under 'exhibition/mostre'.
Sofia	1.9	Process the cheapest plays in the Teatro di Villazzano in November	Not all the plays have their price metadata extracted. Therefore, return the top 5 shows having the minimum prices in a specified theatre or a place chosen.
Sofia	1.10	Process the dance and performances whose cast include Michela Murgia.	Return the records if the director, music composer, conductor, or production attributes contain the name provided, 'Michela Murgia'.
Sofia	1.11	Process the list of movies shown on Friday night in Trento.	A list of movies shown between 19:00 o'clock and 24:00 o'clock on Friday in Trento theater, ranked by price.
Sofia	1.12	Process the locations and availability of the markets where local delicacies and handicraft products are presented.	Extract all the events with types belonging to 'mercati', in combination with market entities to be filtered for the handcrafted items sold there.
Sofia	1.13	Process the expositions to be arranged in Le Gallerie di Piedicastello next month.	Return all the details, including title, descriptions, cost, contact information, subject matters, regarding the exhibitions/mostre to be displayed in the given building structure for the following month.

Sofia	1.14	Process the locations of the buildings hosting scientific meetings.	Derive the geographical locations of structures with respect to events whose subject matters are mapped to one of: archeology, astrophysics, scientific conferences or gadget expositions.
Sofia	1.15	Process a pair of events which can be attended successively in one day.	Although most of the events are held in Trento, there are several communes to be taken into account. Thus, excluding some special events like markets, the procedure can derive combinations of dual events, whose periods are not overlapped.
Sofia	1.16	Process the distributions of theatrical performance types being arranged in the following 3 months.	Cluster the events with respect to each of three following months, calculating the frequencies of each event type for every month, and return the density bars for each cluster.
Sofia	1.17	Process all the painting demonstrations concerned with contemporary and modern arts.	Judging from the apparent absence of critical entries in the dataset, the abstract or description fields must be closely monitored on whether they contain the keywords and their synonyms.
Sofia	1.18	Process distributions of average costs of performances over host buildings and retrieve the buildings whose mean costs are under 20 euro.	Return the building info, the average cost and its standard deviation for every grouped structured having mean prices under 20 euro.
Tojo	2.1	Process all the buildings within the vicinity of 500-metre radius.	Upon extracting the current location, find and return all the hosting construction information with Euclidean displacement lower than 500 metres.
Tojo	2.2	Process frequencies of events according to buildings and weeks in the next month.	Calculate the joint event-counts of a particular hosting structure in a given week of all the 4 weeks of the following month.
Tojo	2.3	Process the events with thematic background revolving around war, sorrow and despair.	Prepare a set of synonymous terms reflecting sorrow and despair and find demonstrations with descriptions accommodating any of those labels.
Tojo	2.4	Process and sort the number of points to visit in 6-hour tour for each of the guided events.	Filtering out the events whose duration last lesser than 360 minutes, the system would count the points of visit in the route for every tour, and return the sorted list of the events according to the frequencies.

Tojo	2.5	Process and identify all the guided tours that contain museums and order by the times different types of museums appear.	From the sequence of routes along the tour, number of types of museums to be visited are counted, weighting each guided tour event to be ordered in a descending way.
Tojo	2.6	Process all the guided tour events with sunny weather and above 10 degree celcius.	Depending on the seasons, the system could adjust degrees differently, giving only the dates whose weather is above that certain threshold, with their respective tour specifications.
Tojo	2.7	Process guided tours leading to the mountainous areas and transpiring within not more than 4 hours.	Assuming there is no obvious label such as 'mountain', it would be better to search along the mountain name list we prepare and roughly match the destination or the intermediary stops.
Tojo	2.8	Process transport means between the selected concert and the festival.	Extract ending time of the first and the starting time of the second after which one is the first play, find the identifications of buses to reach the nearest bus location to the destination of the second event.
Tojo	2.9	Process and order the tourist sites residing at the mountains according to their rankings.	Upon discovery of the sites at the mountains, extract their ratings and number of reviewers who voted and return them with the rating variances.
Tojo	2.10	Process feasible transport means and return them in combination with their prices and expected travel duration.	Assuming the system can manage to pinpoint the nearest bus or train stations to the destination, then it must return the transport vehicles with their ids, prices, and duration time for travel in an ascending order.
Davide	3.1	Give the list of hiking events within Trento.	Returns a list all hiking events by listing the mountain and dates and time
Davide	3.2	Give the list of Hiking events that take place over the weekends.	Retrieves the events under category hiking and date equal to Saturday or Sunday.
Davide	3.3	Give the list of cultural Exhibitions in Trento in the month of December	Returns a list of event category Exhibitions by name, venue and date for the month of December
Davide	3.4	Give the list of bus stations nearest to a given Christmas Market to my residence	Suggests the possible bus stations by station "name", "bus number" and "time" for the next bus
Davide	3.5	Give the list of mountain biking events within Trentino	Retrives all the events with type "mountain biking".

Davide	3.6	Give me the list of all mountains within Trentino.	Retrieves all the mountains only within Trentino that are not so high and more convenient for hiking.
Davide	3.7	Give a list of family festivals in the month of December	Retrieves all the events with type family festival from the list of family events.
Davide	3.8	Give a list of family festivals in the month of December	Retrieves all the events with type family festival from the list of family events.
Davide	3.9	Give a list of all events in Trentino in the month of December	Retrieves all the events with type family festival from the list of family events.
Antonio	4.1	List all the Christmas markets in Trentino	Displays all the Christmas markets by returning attributes “name” , “date” and “location”.
Antonio	4.2	Give a list of Christmas markets in Rovereto that take place over the weekend	Displays the attributes “name” and “location” of only the Christmas markets that take place on Sunday and Saturday
Antonio	4.3	Give the closest Christmas market from my residence	Suggests the nearest Christmas markets depending on the location of the user by listing their name and “location”, “date” and “duration”.
Antonio	4.4	Give the list of Christmas markets that operate beyond midday	The system retrieves only the markets with duration greater than 12:00 o clock
Antonio	4.5	Give the list of other weekly markets within Trento?	Display all the markets that take place every week coupled with their location and duration
Antonio	4.6	Locate the market running on Thursdays?	The system will show a map directing them to the market on that particular day.
Antonio	4.7	Give the list of cultural festivals within Trentino ?	A list of cultural festivals will be displayed which will include the regions they are taking place, the date and time.
Antonio	4.8	Give the list of theatrical events within Trento	Retrieves a list of concerts, movie nights , according to their titles , venue and time
Antonio	4.9	Give a list of movies that do not take more than 2 hours to watch	Retrieves all movies whose duration is less than 2 hours to watch by movie title.
Antonio	5.0	Give a list of movies that are translated to English	Retrieves all the list of event type movies language is both Italian and English
Antonio	5.1	Give a list of family events that take place over the weekends	Retrieves all the list of events with type Family that are to occur either on Sunday of Saturday
Antonio	5.2	Give list of available public transport means to the nearest Family event	Retrieves train, bus stations from a given point to the nearest family event.

NUM	TYPES	PROPERTIES

1:1-2-3-4-5	Market	type
1:11-12-13-14-15-16-17-18	Movie	genre, price
1:1-2-3-4-5-9-11-17-18, 2:1-2-3-7	Address	province, city, village, street, number, CAP
1:1-2-3-4-5-6-7-8-11 2:5-6	Period	starting date, ending date
1:4-14, 2:8	Duration	duration per day
1:6-7-8-9-10	Performance	type, price
1:10-16-18	Actor/Actress/director	name
2:2	Shop	type, offer
2:1-5-7-8-9	Exhibition	type, content
2:3-4-6	Cycle Route	type

2.2.2 Initial Datasets description

Since we are focusing on the events being organised in Trento, we decided to work with the local datasets. To retrieve the local events we have chosen the website www.trentino.com, listing various events like 'Farmer's markets', 'Festivals', 'Guided tours' etc. being held in the Trento area on a given date. The listed events contain information on the dates, timings and location with exact longitudes and latitudes of the event with the relevant contact information. Another important category of tourist event is cinema. So, we utilized the website www.crushsite.it. It is an on-line agenda of culture, art and entertainment . It contains the events organized in Trentino and the main initiatives of Alto Adige , some sector news , the files of over ninety regional cultural subjects involved in the project. The website provides details of the cinema being released in the local theatres with the details on the movie's cast, director, language, genre, duration in minutes and the most important address of the theatre and the ticket booking information. To gather more data on the evnets organized in Trento we used another website www.visittrentino.info which presents a list of events with their start and end dates, timings, address, type of the events, organizer's contact information, location in lat/long etc.

2.2.3 Datasets collection process

A combination of methods such as web-scraping and searching for the public datasets has been employed in order to prepare the unrefined initial data materials. First, the most prominent datasets including *Events in Comune of Trento* and *FamilyAppointment* were found on the Trentino website for the public data [1]. Secondly, in an effort to figure out what websites could provide datasets with a sufficient relevancy and a broad spectrum of properties, we have listed which sources Google references to integrating the event data, narrowing down to Viator [2], Facebook [3], TrentoToday [4] and Trentino [5]. Subsequently, we have used BeautifulSoup [6] to prepare events concerning the flea market from Trentino website, whereas the Scrapy [7] has been use of for extracting the datasets with respect to outdoor activities and meetings from Viator and Trentotoday platforms. Finally, the facebook platform selectively recommends events, mostly online, according to the particular taste of a user, which we took advantage of, by trying

[1] http://www.comune.trento.it/api/opendata/v2/content/search?classes=event
[2] https://www.viator.com/en-GB/Trento-tours/Outdoor-Activities/d29399-g9?sortType=featured
[3] https://www.facebook.com
[4] https://www.trentotoday.it/eventi/tipo/incontri/dal/2020-03-01/al/2020-12-31/
[5] https://www.trentino.com/en/events/results/
[6] https://pypi.org/project/beautifulsoup4/
[7] https://scrapy.org/

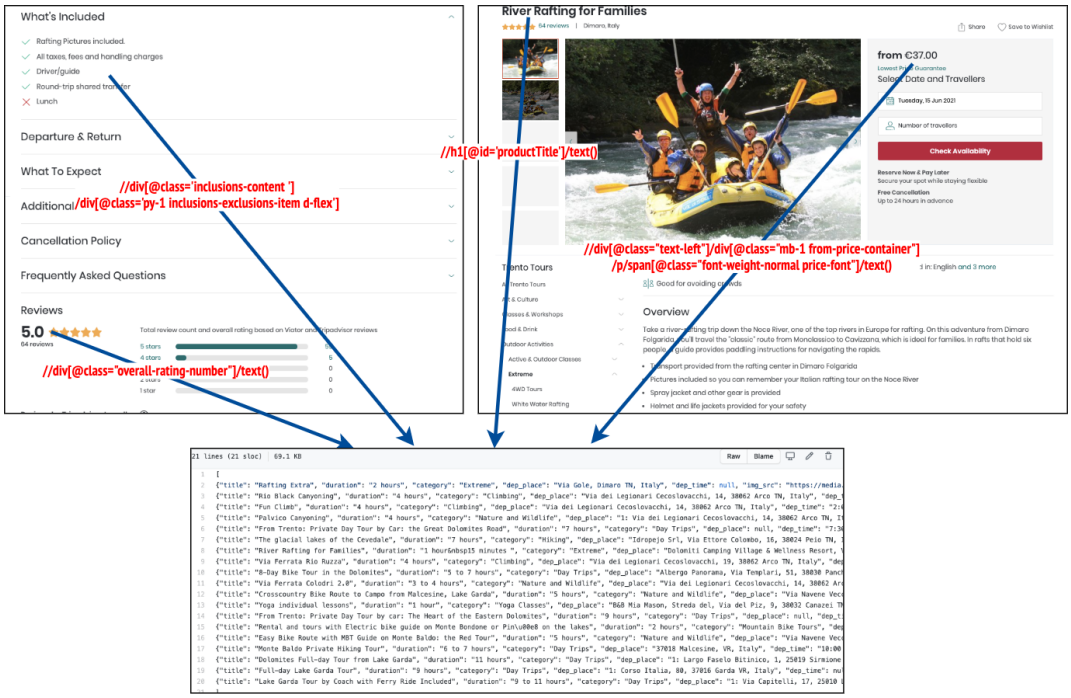


Figure 1: An illustration of utilizing XPath to scrape dataset properties from an exemplary website (e.g. Viator)

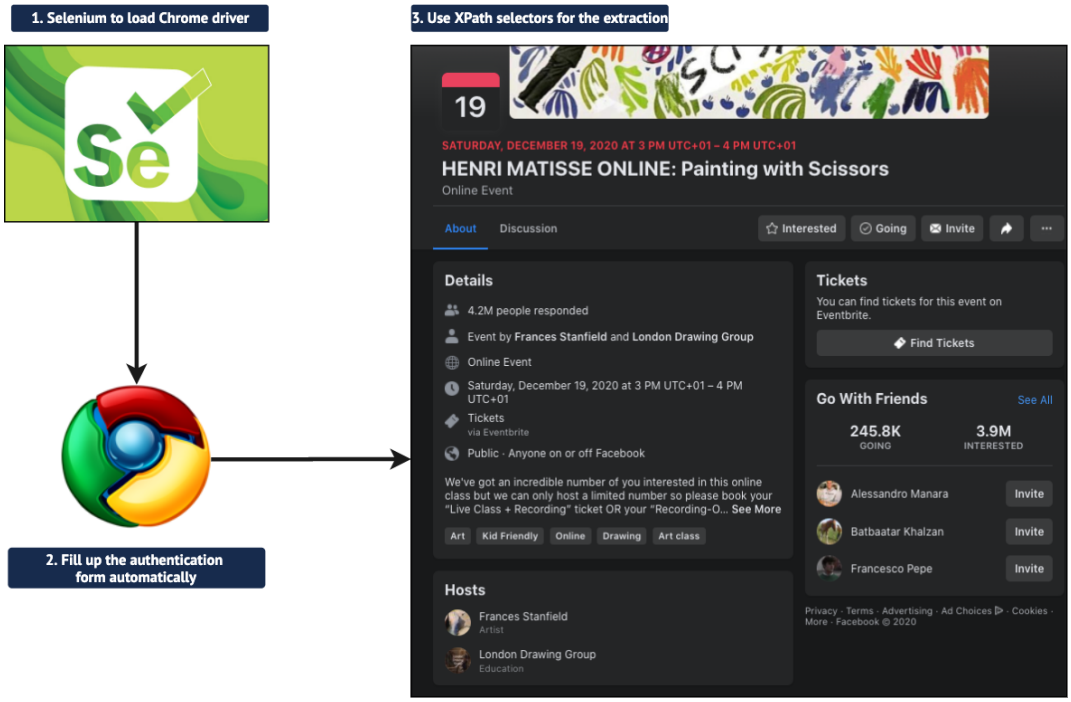


Figure 2: An application of Selenium for authentication passing to facilitate scraping the data features from Facebook

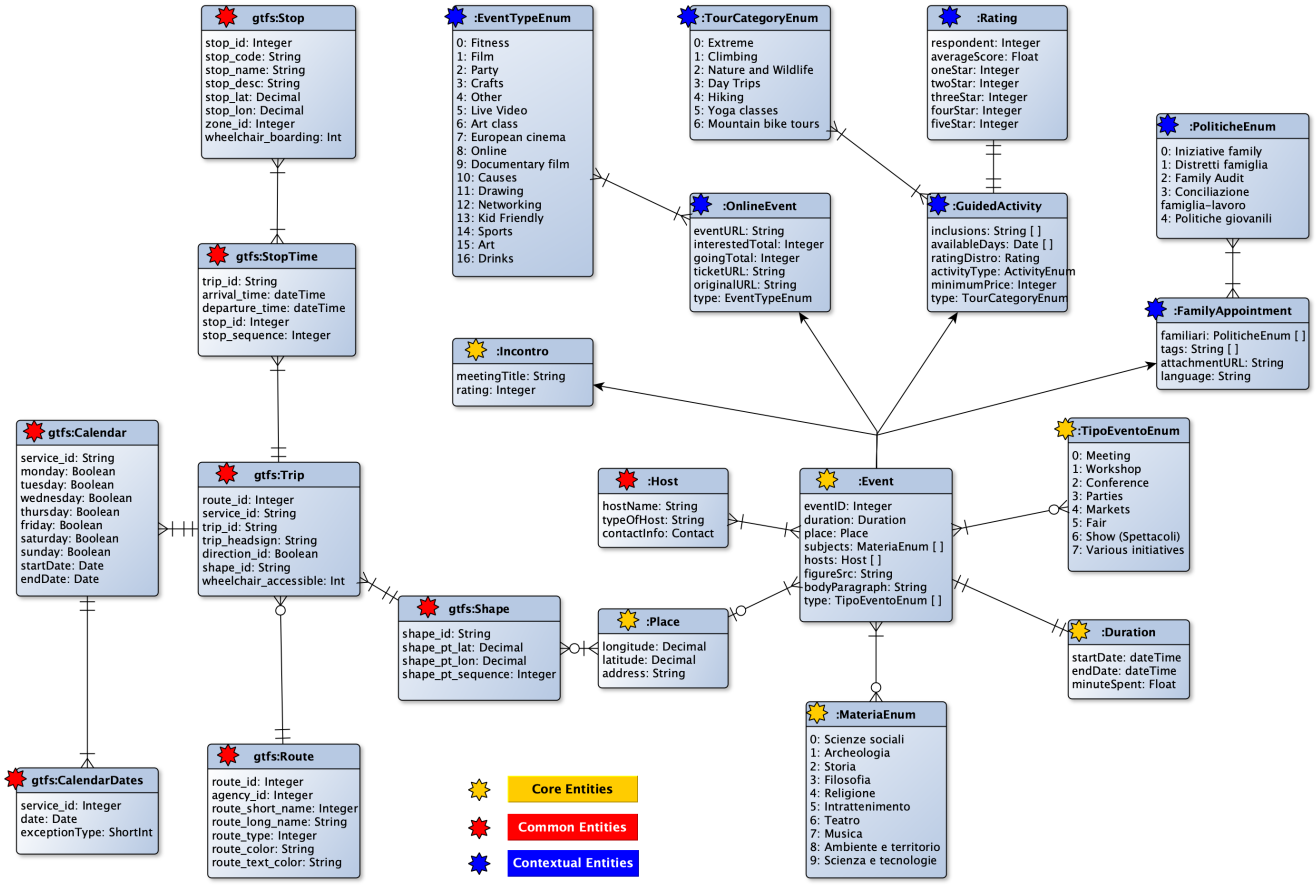


Figure 3: Extended Entity Relationship Diagram depicting the hierarchical relations and conjunctions of different levels of entity types

to extract the event data properties given. The plan was to use the Selenium [8] to load a Chrome driver to be under our supervision so as for us to authenticate into Facebook programmatically and iterate through various events.

2.2.4 Inception level evaluation

As a result of having comparisons between the datasets currently at our disposal that we have scraped and the competency query sets, we have estimated that the coverage is around 0.23% and the flexibility 76.31%.

2.3 Informal Modeling

According to the compatibility of the requirements from the queries defined earlier with the entity types which can be formed out of the available datasets, a rough blueprint for the application frame is constructed as follows.
As illustrated in the entity relationship diagram (fig. 3), the main application is supposed both to (1) deliver requested events belonging to various typologies and (2) to promote transportation means and routes for carrying the customer between a pair of events. Nevertheless, the latter is not strictly followed as there would be many instances where the origination or the destination cannot be the geographical information with respect to an event, but randomly given points (e.g the current location). In addition, the uniqueness for this modelling prompts us to

[8] https://www.selenium.dev/

deal with a great deal of temporal and spatial relationships between the transit system and events being transpired regardless of any other intervention, emphasizing the significant essences that must be optimized throughout the rest of the phases.

2.3.1 Etypes and Properties

1. Criteria in creation of subclasses

Decomposing an entity type further into subcategories in presence of semantic heterogeneity (in our case, it's "Event") from different sources can lead to two potential issues.

First, failure to provide the set of defining properties for a subcategory might obscure the significance of branching out a new category. Exemplifying from our project, "Market" etype missing the properties of items to be sold there and the others highlighting its distinction from surrounding etypes puts itself in a position where "Market" is a regular "Event". Secondly, the datasets springing from completely incongruent platforms establish inherently divergent entity types, while having conceptually synonymous notions. For instance, the facebook-dependent properties such as 'interested' and 'going' contextualize its own entities virtually nowhere around the rating systems which could be offered from different websites. By taking into account these two points, our solution attempted to distinguish conceptual hierarchies as much apart as possible with respect to their difference in properties and conceptually abstract definitions.

2. Composition and Arrangement

In this subsection, there is a comprehensive explanation concerning the entity classification into common, core and contextual data models. In order to designate etypes in their correct environment, we might want to keep in mind two things: rigorous statements corresponding to each category and the ultimate goal of our application. The assertions clearly reflect on our understanding of the entity divisive system, meaning they merely are assumptions being hold true for the remaining parts.

- **Core etypes:** fundamental building blocks (denoted under yellow stamps) without which the main application cannot function in accordance with its primary requirements. Event is the most generic yet the most instrumental entity in the program, defined as taking place in a certain location in a given duration. Although there are several different types of sub-events to emerge, it is efficient to have them as enumerated types (TipoEventoEnum and MateriaEnum), assuming the rarity of data records belonging to each type and the lack of diversity having them distinguished. From there, GuidedActivity, which compels participants to display physical endeavors and initiatives, branch out to create a subclass. Due to the source web from which it has been extracted, it also incorporates an extensive list of local properties.

- **Common etypes:** non-defining entities that could be found in any other types of applications, yet contributes to making up the integral system (denoted under red stamps). Shape, a sequence of geographical points demonstrating a route for a vehicle to race along, has its points represented by Place, allowing the query messages to flow across from the core entities. As long as a system gets hold of the correct shape, it can capture its Trip, an instance of a vehicle running over the shape, StopTime, a temporal information about the trip, and the literal Stop for the destination. Plus, with the help of Calendar and CalendarDate, it enables for users to run relatively advanced queries to see the accessibility of the chosen vehicles.

- **Contextual etypes:** phenomenal entity types that would enhance the accuracies and specificities of the query answering process in regards with the core etypes. In terms of event, we divided the generic concept

Figure 4: Results prompted from LOV in response to the search term 'Event'

into Online, Family and Guided events. Each has been born maximizing factors described earlier in the section, the platform-wise uniqueness and the conceptual uniqueness.

2.4 Formal Modeling

2.4.1 Namespaces for the Domain

Our transition to the formalization phase is marked by our undertaking in search of already existing vocabulary terminologies that would cover the clumsy etypes defined in the informal modelling phase and the namespaces containing them. This is in an attempt to bring, as closer as we could, the gap between our rough sketch for the event domain and the commonly acknowledged standard for the same domain, so that we could get stuck by a better idea to improve the current model.

The most frequent term 'Event' in the EER at our disposal possesses relatively a fewer number of senses, ruling out the most of our search space in Linked Open Vocabularies [9].

As demonstrated in the search result capture, there are two possible formal schemas for us to appropriate into our application (fig. 4).

- At first glance, schema.org [10] appears to fit the profile we're looking for the most parts because it is a

[9] https://lov.linkeddata.es/dataset/lov/
[10] https://schema.org/Event

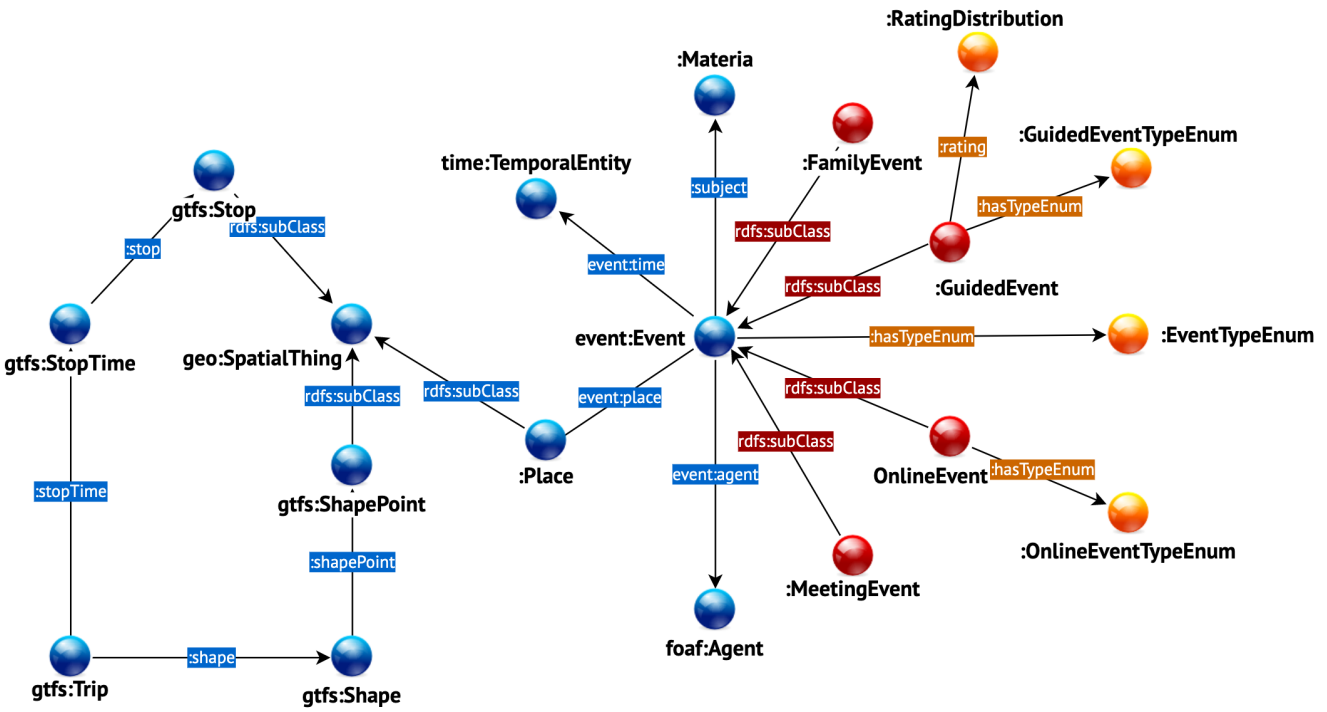


Figure 5: Schema Knowledge Graph formalized from the informal settings

widespread shared terminology-space and the predicates belonging there are able to cover the majority of our object (and data) properties, which means our data integration to a bigger knowledge graphs could be guarenteed and it increases the chance that concepts and senses in the UKC-KB are located way more accurately if we start from this namespace. Unfortunately, the problem arises when it comes to the sub-event categorization. It is true that we enumerated our event types instead of creating subcategories for each, just like schema:Event offers, such as detaching off 'TheaterEvent', 'MusicEvent' and 'DanceEvent'. However, as in the argument from the previous phase, the scarcity of mining tools for extracting essential and defining properties for each subtype led to the decision to leave them under the same entity type. For example, let us say, we wish to derive 'producers', 'cast', 'stage designer' from the description (e.g. body paragraph) property of 'TheaterEvent' entity. Not only they are written in human natural language, but also they vary in ideosyncrasies authors display them in, making it impossible for us to write regex expressions to filter them out. In addition, although we could use APIs for named entity recognition, these cannot manage to contextualize what they have captured even if the entity in the description and its type have been recollected.

- On the other hand, event [11] is a lightweight ontology that incorporates foaf [12], time [13] and geo [14] to represent its key connections. Given our circumstances, assuming the unorthodox semantic representation we have of the touristic event domain, it is a better choice to embrace this ontology providing us a customization flexibility with respect to dividing the parent class into unfamiliar application-specific subevents. Though it is in no way the best schematization, considering the project period, lack of data diversity and the amount of effort necessary to fragment the properties into smaller ones as described above, we assume our assertions up

[11] http://motools.sourceforge.net/event/event.html
[12] http://xmlns.com/foaf/0.1/
[13] http://www.w3.org/2006/time
[14] http://www.w3.org/2003/01/geo/wgs84$_{p}os$

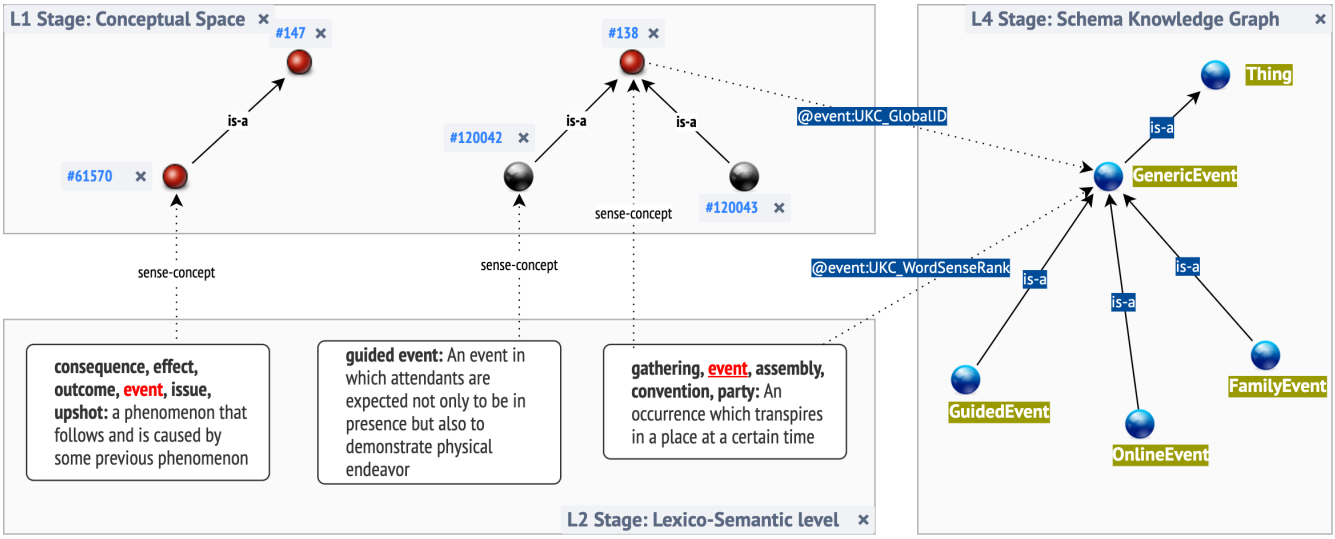


Figure 6: Annotations with reference to the UKC

to the point provided sufficient rational causalities in our decision-making.

- Here (fig. 5), the important change with respect to connecting the transportation and the casual event parts stream along the etype, Place, modified to become a child concept of the broader term, geo:SpatialThing. What emphasizes it from the previous phase is that SpatialThing is a container for the lat/long properties, a common signature for every type related to the geographical point. Place itself diverges as it has a proper address where event to be organized, whereas ShapePoint has the sequence to identify which point it is for a certain Shape. Eventually, this custom arrangement allows us to find the most appropriate trip to the event place, by locating the nearest shape (almost a directed route), then tagging along with it to find the available time and stations to reach.

- Finally, another substantial part of our application, the data representation regarding transit feed system, is covered by gtfs [15] ontology. Given the datasets, we are sure to decorate them with the classes provided by the namespace. Nonetheless, it doesn't put forward the object properties connecting the domains and ranges, apparently due to changeability in their designs.

2.4.2 Annotations with UKC

In spite of the original instruction for us to decorate the etypes using annotation properties UKC_GlobalID and UKC_WordSenseRank, we decided to follow the etype-GID_ID format referred in the demo presentation, in an attempt to avoid potential naming violation that we could encounter in the later stages.
With respect to the concept availability, there appeared to exist the entity types concerning public transportation sub-system, perhaps adjusted by the other teams, offering resources for us to cut our updating senses in half. Subsequently, we prepared the following five major concepts as well as their hierarchical relationship metadata:

- **OnlineEvent [GID-120042]** IS-A Event [GID-138]: A special type of event with the participants prompted to engage in remotely (Word form: Remote Event)

[15] http://vocab.gtfs.org/gtfs.ttl

- **FamilyEvent [GID-120044]** IS-A Event [GID-138]: A type of event organized by a local volunteers to enrich communication exchange between families (Word form: Family Appointment)
- **GuidedEvent [GID-120043]** IS-A Event [GID-138]: A particular arrangement in which attendants are expected not only to be in presence but also to demonstrate physical endeavor (Word form: Guided Outdoor Activity)
- **OnlineEventType [GID-120045]** IS-A Enumeration [GID-34789]: Enumerated types for the online events, including such as painting, sightseeing and visiting museums (Word form: Remote Event Category)
- **GuidedEventType [GID-120046]** IS-A Enumeration [GID-34789]: Categories of physical activities, especially mountain hiking, skiing and climbing (Word form: Guided Outdoor Activity Typology)

2.4.3 Data level

As in the previous phase the data level section here, reports the description of the new version of the datasets, after formatting operations.

2.4.3.1 Formal Modeling datasets management

In this section are reported the operations and the tools adopted to format the dataset collected, in order to align them to the ontology definitions generated at schema level.

2.4.3.2 Datasets metadata documentation

In this section eventually new metadata information are added in order to describe the evolution of the datasets.

2.4.3.3 Variance respect Informal Modeling datasets

This section aims to define the variance between the data elements (datasets and attributes within them) produced in this phase, and the initial datasets collected in the previous phase. This a way to define the quality of the outcomes for the current phase as well as the alignment of the overall project development process.

2.5 Data integration

This section is dedicated to the Data Integration phase description.

2.5.1 Data integration operations and tool

This section is dedicated to the description of the usage of the data integration tool that allows to map the datasets generated and well formatted in the previous phases, with the final ontology generated. The last datasets adaptation performed using the tool, as well as the mapping operation are here detailed.

2.5.2 Variance respect Formal Modeling datasets

The last section of the data integration phase aims to describe the variance, analyzing the differences, between the datasets integrated with the ontology, in the data integration platform which contain the KG, and the datasets collected in the previous phase. This analysis can highlight the results of the operations performed during the final phase of the data integration process.