Linking Smart Cities Datasets with Human Computation – the case of UrbanMatch

Irene Celino[1], Simone Contessa[1], Marta Corubolo[2], Daniele Dell'Aglio[1], Emanuele Della Valle[2,1], Stefano Fumeo[1], and Thorsten Krüger[3]

[1] CEFRIEL – Politecnico di Milano, Milano, Italy – {name.surname}@cefriel.it
[2] Politecnico di Milano, Milano, Italy – {name.surname}@polimi.it
[3] Siemens, Corporate Technology, Munich, Germany – thorsten.krueger@siemens.com

Abstract. To realize the Smart Cities vision, applications can leverage the large availability of open datasets related to urban environments. Those datasets need to be integrated, but it is often hard to automatically achieve a high-quality interlinkage. Human Computation approaches can be employed to solve such a task where machines are ineffective. We argue that in this case not only people's background knowledge is useful to solve the task, but also people's physical presence and direct experience can be successfully exploited. In this paper we present UrbanMatch, a Game with a Purpose for players in mobility aimed at validating links between points of interest and their photos; we discuss the design choices and we show the high throughput and accuracy achieved in the interlinking task.

1 Introduction

Cities are defined smart when their investments in the human and social capital, as well as in the communication infrastructures are aimed at fuelling a sustainable economic growth and a high quality of life [6]. Specifically, current research investigates the impact of ICT on the development and improvement of smart cities with respect to several dimensions, from people to government, from mobility to environment, etc. In this context, a key to realize smart cities is to involve smart citizens by raising their awareness, participation and contribution.

Big industrial players are focusing their research and innovation around smart cities; some examples are the initiatives carried out by Siemens[4], IBM[5] and CISCO[6]. Public authorities are also becoming more and more attentive to adapt their political agenda to fulfil this smart cities vision, in particular through an open data strategy.

Geo-spatial data and information related to entities located in the physical world are among the first sources that are published openly – and often also freely – on the Web; valuable examples are Ordnance Survey location data in the UK[7], GeoLinkedData.es in Spain[8], GeoNames geographical database[9] and

[4] `http://www.usa.siemens.com/sustainable-cities/`
[5] `http://www.ibm.com/uk/smarterplanet`
[6] `http://www.cisco.com/web/strategy/smart_connected_communities.html`
[7] `http://www.ordnancesurvey.co.uk/`
[8] `http://geo.linkeddata.es/`
[9] `http://geonames.org/`

the community-driven OpenStreetMap[10]. The Semantic Web community also has showed interest in geo-spatial data: OpenStreetMap was turned into Linked Data by the LinkedGeoData project [26] and the Open Geospatial Consortium is standardizing GeoSPARQL[11], a spatial extension of the SPARQL language.

For the last years, we have been experimenting with geo-spatial data – especially with those related to urban environments – in order to build Linked Data-enhanced applications and services. The used datasets and the applications objectives were diverse: points of interest and event data to plan journeys [9]; traffic sensors data and road topography to predict the most suitable path [10]; urban regulations to update road sign information [17]; social media to provide location-based recommendations of restaurants [3].

While the large availability of urban data is an advantage in realizing such kind of services, the poor quality or the doubtful trustworthiness of the information source strongly hamper a large-scale adoption of those data. Imprecise or outdated information, sparse or heterogeneous distribution of data are just some examples of the obstacles to a proper reuse of geo-spatial (linked) data. Our experience tells that inconsistencies and imprecise data can be detected – and their quality improved – by a small amount of manual work that does not require specific skills, but often the physical presence in the urban environment [17].

Human Computation [29] is the paradigm to leverage human capabilities to solve tasks that computers are not yet able to properly undertake. A Human Computation approach is often employed to solve data quality tasks.

Our research question can be formulated as follows: is it possible to exploit people's *physical presence* in the environment to improve geo-spatial data quality? Can we build a new generation of Human Computation techniques based on the contributors' *direct experience* (instead of a specific domain expertise)?

To check our hypothesis, we built UrbanMatch [7], a location-based Game with a Purpose [28] in the form of a mobile application[12]. Specifically, UrbanMatch is aimed at exploiting players' experience of the urban environment to correctly link points of interests in the city with their most representative photos retrieved from Web sources. The paper's *contribution* lies in the modelling of the POI-photo linking as a record linkage problem and the realization of the game using this formalization; we also experimentally determined the best combination of the model's parameters, in order to optimize the trade-off between the number of links created per playing hour (system throughput) and the accuracy of those links.

The remainder of this paper is organized as follows. Section 2 introduces the related work; Section 3 defines the problem statement, while the process to achieve the game purpose is detailed in Section 4. Section 5 explains the mechanics of the UrbanMatch game, while the evaluation is illustrated in Section 6; finally Section 7 draws some conclusions and future work.

[10] `http://www.openstreetmap.org/`
[11] `http://www.opengeospatial.org/standards/requests/80`
[12] UrbanMatch is available on iTunes app store at `http://bit.ly/um-itunes`.

2 Background and Related Work

Our work focuses on user interaction for link elicitation and validation for Linked Data in urban scenarios. It is centred on Linked Data and it is based on the results of three research areas: data linking, data quality, and human computation.

2.1 Data Linking and Linked Data

Data Linking is the problem of deciding whether resources belonging to different data sources are referring to the same entity. It is rooted in the record linkage problem studied in the databases community since the 1960s [13, 20, 32].

Record linkage is a challenging task, as deciding if records match is often computationally expensive and application specific [5]. The former is because a combination of string similarity algorithms have to be used, the latter because it is difficult to provide a general solution which works well with heterogeneous datasets. For instance, the techniques used in linking scientific datasets will be different from the ones used for linking CRM datasets.

In this paper, we are particularly interested in referring to the formal definition of record linkage introduced by Felligi and Sunter in [13], which we use in the rest of the paper. When linking the records in two databases A and B, the idea is a) to classify links in the comparison space $\Gamma = A \times B$ into M – the set of matches –, and U – the set of non-matches –; b) to compute for each link γ a score s as ratio of probabilities $P(\gamma \in \Gamma|M)/P(\gamma \in \Gamma|U)$; and c) to use the score s to divide the comparison space in three disjoint sets using an $UPPER$ thresholds, a $LOWER$ threshold and a decision rule. If $s > UPPER$, then the link is correct; if $LOWER \leq s \leq UPPER$, then the link needs to be assessed by an expert; if $s < LOWER$, then the link is incorrect.

Establishing links between datasets published as linked data [27] is a problem rooted in record linkage, but can benefit from the availability of ontologies describing the datasets to be linked, and, thus, from existing ontology matching solutions [22, 12]. At the time of writing, SERIMI [2], Zhishi.links [21] and AgreementMaker [8] are the best data linking solutions emerged from the Data Interlinking track of the OAEI 2011 challenge [11].

2.2 Data Quality and Linked Data

Data Quality [23, 4] is the discipline that studies the most appropriate and relevant features to describe the value of data.

A key point of Data Quality is the context-dependency: given a dataset, its quality can be very high w.r.t. the fulfilment of some tasks but very bad w.r.t. other ones. As pointed out in [16], "the perception of information quality (on the WWW) is highly dependent on the *fitness for use* being relative to the specific task that users have at their hands". In other words, it is not relevant (and not always possible) to define absolute quality factors [19]. More specifically, [23] enumerates the following factors contributing to *fitness-for-use*: accuracy, completeness, consistency with other sources, timeliness, accessibility, relevance, comprehensiveness, easy to *read* and easy to *interpret*. In this work, we evaluate our solution using accuracy and throughput (see Section 2.3) as a proxy for completeness.

It is worth noting that the Linked Data best practices alone [15] assure more quality than "raw data" in "closed" databases because: *a*) data becomes *accessible over the Web* rather than being closed up in silos; *b*) the use of *shared vocabularies* makes the data both easier to "read" (i.e. user information needs can be satisfied by a single SPARQL query instead of requiring many dataset-specific queries) and easier to "interpret" (i.e. shared vocabulary semantics can be used to verify data integrity and/or infer implied data); *c*) the *presence of links* makes it also possible to verify consistency across different sources.

However, the assessment of data quality factors like accuracy, completeness, timeliness, relevance and comprehensiveness of data is intrinsically a hard task that Linked Data best practices do not make any easier. As one can expect, the quality of published Linked Data is variable and the community has started to follow data quality with growing interest. Flemming worked on the definition of quality criteria for linked data sources [14]. She grouped the criteria to describe data sources in four categories: *content* (the quality of the data as available in the dataset), *representation* (an evaluation of the data serialization), *usage* (the measurement of the data "fitness for use") and *system* (indicators about the publishing system).

2.3 Human Computation and Linked Data

As we showed in the two previous sections, data linking and data quality are hard problem for computers and subjective in nature. We, as humans, are perfectly capable of both tasks, but we are not necessarily willing to. Human Computation [29], however, demonstrated that "computations" of this kind can be carried out by groups of people if motivated by the right incentives.

The incentives to make people contribute can be of different kinds: they can give the participant an explicit and concrete reward (like in the popular Amazon Mechanical Turk[13] in which people are paid to perform small and simple tasks) or they provide a different kind of implicit or more abstract return, for example by means of entertainment like in Games with a Purpose [28] (GWAP).

In this paper, we are specifically interested in the design and evaluation of GWAPs as UrbanMatch is a GWAP. Having created many GWAPs (e.g., ESP Game, Peekaboom, Phetch, and Verbosity), Luis Von Ahn and Laura Dabbish reports in [31] on three game-structure templates that generalize successful instances of Human Computation games: input-agreement games, inversion-problem games, and output-agreement games. In input-agreement games, players must determine whether they have been given the same input; in inversion-problem games, given an input, a player produces an output, and another player guesses the input; in output-agreement games, players are given the same input and must agree on an appropriate output. UrbanMatch is an output-agreement game.

[13] `http://mturk.com/`

UrbanMatch is not the first GWAP proposed by the Semantic Web community. GWAPs have been already used to cover the complete Semantic Web life-cycle [25]. A dedicated community portal was recently set up[14] to collect those games. A good showcase is the Linked Data Movie Quiz [1], that builds a cinematographic game based on the available movie-related Linked Data showing that "the answers are out there; and so are the questions".

The metrics [31] proposed to evaluate GWAPS include throughput and average lifetime play (ALP). The *throughput* of a GWAP is defined as the average number of problem instances solved per human hour. The higher the throughput the more effective the GWAP. However, a GWAP with a high throughput that fails to attract and keep players is useless. The *ALP* is a proxy for the intangible enjoyability of the GWAP. It is defined as the overall amount of time the game is played by each player, averaged across all people who have played it. A successful GWAP like the ESP game [30] has a throughput of 233 problem instances solved per human-hour and an ALP of 91 minutes. We use those metrics to evaluate Urban Match in Section 6.

3 Problem Statement

UrbanMatch aims at linking urban related data sets. More specifically, the purpose of UrbanMatch is to derive meaningful links between a datasets containing the points of interest (POIs) in a urban environment and a dataset with the images depicting those POIs and retrieved from Web social media; among all photos taken in the proximity of a POI, UrbanMatch is designed for linking the most representative ones to that POI.

We selected the first dataset A of POIs from OpenStreetMap/LinkedGeoData and we retrieved the second dataset B collecting photos from social media sharing sources, namely Flickr and Wikimedia Commons. The first edition of the UrbanMatch game is released for the city of Milano in Italy, thus the POIs in dataset A are *tourist attractions* in Milano, i.e. entities in LinkedGeoData [26] that are instances of classes like `lgdo:Monument`, `lgdo:Historic` or `lgdo:Landmark`.

Dataset A contains the POIs aggregated by *playable place*. A playable place is an open area (like a square or a park) that is physically adjacent to at least two tourist attractions. Playable places are retrieved via a spatial-enhanced query on OpenStreetMap. Given the list of playable places with the corresponding POIs, we appealed to an expert judgement (a person, among the authors, familiar with the city and its notable locations) to filter out the irrelevant elements and to complete a list of alternative names/labels to indicate the selected POIs. The result was a set of 14 playable places with 34 POIs in Milano.

To retrieve the dataset B of the photos, we used the POIs geographic coordinates to perform a *spatial query* on the image sources – Flickr and Wikimedia Commons – by invoking the respective API. This location-based query was enhanced with other information about the POIs: on Flickr API, geographic coordinates were used together with a *keyword search* by using the alternative POIs

[14] `http://www.semanticgames.org/`

names/labels. On the other hand, Wikimedia Commons – the media database related to Wikipedia – puts in relation its photos with the Wikipedia page that describes the depicted POI; thus, the retrieval requests mix the geographic coordinates with a *"conceptual" search*, comparing a Wikipedia page with the "concept" of the respective POI. The result of the photo selection was a set of 11,287 photos of Milano POIs.

We can formulate the *data interlinking problem* that UrbanMatch aims to solve as a record linkage problem using the formal definition of record linkage introduced by Felligi and Sunter in [13] (already cited in Section 2.1). Thus, we define the set Γ of all possible links between POIs and photos as the comparison space[15] between the two datasets, i.e. $\Gamma = A \times B$. Each link $\gamma_{p,n} \in \Gamma$ can be seen as an RDF triple of the form:

```
<POI-n> foaf:depiction <photo-p> .
```

in which `<POI-n>` is the URI of the POI n in dataset A and `<photo-p>` is the URI of the photo p in dataset B.

Data interlinking is achieved when all the links γ in the comparison space are classified in two sets: the set M of "matches", i.e. of *correct links*, and the set U of "non-matches", i.e. of *incorrect links*. Each link $\gamma_{p,n}$ is associated with a score $s_{p,n} \in [0..1]$ that represents the probability of the link to be correct; two thresholds are usually defined – $UPPER$ and $LOWER$ – so that:

$$\text{if } s_{p,n} > \text{UPPER} \quad \text{then } \gamma_{p,n} \in M$$

$$\text{if } s_{p,n} < \text{LOWER} \quad \text{then } \gamma_{p,n} \in U$$

The comparison space Γ between two datasets A and B can be seen as divided into three disjoint sets: M, U and the set C of unclassified links, for which:

$$\text{LOWER } \leq s_{p,n} \leq \text{UPPER}$$

Solving the data interlinking problem, therefore, requires the ability to alter the score of the links $\gamma_{p,n} \in C$. Those links represent our *candidate links* that need to be assessed to be classified either in M or in U.

In the case of UrbanMatch, the candidate links in C are those links whose quality is not appropriate according to the *fitness-for-use* principle [23]. For example, a candidate link could connect a POI with a photo that frames the inside of that POI, or a non-evocative detail of the POI, or people in front of the POI. The central idea of UrbanMatch is to use a GWAP to ask players to assess the candidate links in C and to alter the score of each link until it falls either in M or in U.

Each link γ between a POI in the dataset A and a photo in the dataset B is given an initial score s between $LOWER$ and $UPPER$, so that all links initially belong to the subset C of candidate links. To bootstrap the UrbanMatch approach, we manually modified the score of some candidate links in C to a value either greater than $UPPER$ or lower than $LOWER$. To this end, we appealed

[15] A reduction of the comparison space Γ by partitioning is explained in Section 4.

to an expert judgement to select some photos depicting the POIs: on the basis of this selection, some links moved from C to M (because the photos depicts *for sure* the POIs) while some other links moved from C to U (because they do *not* depict the POIs). The result of this preparation phase was a comparison space Γ with 196 links in M, 382 links in U and 37,413 link in C. Table 1 recaps the initial dataset with the detail for each playable place in Milano.

Table 1. The initial dataset of UrbanMatch Milano (created in spring 2012).

Playable Place Name	bootstrapped photos	total photos	POIs	links in M	links in U	links in C
Piazza dei mercanti	16	301	2	16	16	634
Piazza Sant'Ambrogio	12	180	2	10	10	380
Piazza del Duomo	32	3,884	5	32	128	19,580
Piazza Duca d'Aosta	11	1,032	2	11	11	2,086
Via Legnano	10	418	2	10	10	856
Via Conservatorio	13	325	2	13	13	676
Viale Alemagna	12	217	2	12	12	458
Largo Marco Biagi	9	973	2	9	9	1,964
Corso di Porta Ticinese (1)	12	142	2	12	12	308
Corso di Porta Ticinese (2)	11	376	2	11	11	774
Via Gioia	15	110	2	15	15	250
Corso Venezia	10	979	2	10	10	1,978
Piazza della Scala	20	885	4	20	80	3,620
Piazza Cairoli & Viale Petofi	15	1,269	3	15	45	3,852
Total	198	11,089	34	196	382	37,413

4 Achieving the UrbanMatch Purpose

The UrbanMatch game was designed to let the players rate the candidate links. However, presenting players directly with the RDF links is not a user-friendly way to let them solve the task. Moreover, if the players are not in the urban space or if they do not have enough background knowledge about the POIs, it could also be difficult for them to say, for example, if a photo actually depicts "Palazzo della Ragione". For those reasons, UrbanMatch is designed as a single-player mobile game to be played on the go, in which the players are presented only with the photos and are asked to pair those that represent the same POI around them. The players may not know the name of the depicted POI, but if the photo is representative, they can recognize it.

The photo pairs selected by the player are used by UrbanMatch to alter the scores of the candidate links between those photos and the POIs around the player. Let us assume that UrbanMatch shows the player two photos a and b and it wants to assess if the two photos are both linked to the same POI 1. Let us also assume that the link $\gamma_{a,1} \in M$, i.e. the link between a and POI 1 is correct, and that $\gamma_{b,1} \in C$, i.e. the link between b and POI 1 is candidate. The player can decide whether to pair the two photos or not. If the player pairs the two photos, the score of the link $\gamma_{b,1}$ is modified; the new value of the score $s'_{b,1}$ of the link $\gamma_{b,1}$ between the photo b and the POI 1 is increased using the formula in Equation 1

$$s'_{b,1} = s_{b,1} + K_{pos} \tag{1}$$

where K_{pos} is a positive constant that counts for the *positive evidence* provided by the player.

Otherwise, if the player does not pair the two photos, the new value of the score $s'_{b,1}$ of the link $\gamma_{b,1}$ between the photo b and the POI 1 is decreased using the formula in Equation 2

$$s'_{b,1} = s_{b,1} - K_{neg} \tag{2}$$

where K_{neg} is a positive constant that counts for the *negative evidence* provided by the player. Collecting positive and negative evidences for each link in C, UrbanMatch alters the score of each candidate link until it is categorized as belonging either to M or to U.

An important issue arises when modifying the links' scores: are players reliable? We can certainly trust a large majority of the players to play earnestly, but we need to consider that some players can cheat or misunderstand the task, thus giving wrong answers. As proposed in [31], we can mitigate the risk of trusting erroneous inputs with two strategies: *i*) repeating the same task multiple times to randomly picked users, and *ii*) testing the player reliability.

The approach described in Equations 1 and 2 is ready to support the first strategy, by opportunely tuning the values of K_{pos} and K_{neg}. As noted in [31], this strategy can guarantee the correct assessment of link quality with arbitrarily high probability.

The second strategy can be embedded in Equations 1 and 2 by testing players multiple times per game and evaluating their reliability on the basis of the number of errors they make. As a result of the bootstrapping, a number of incorrect links exist; UrbanMatch puts them in the game as verification cases. For example, let us assume that we have only two POIs (1 and 2), and UrbanMatch shows the player two photos (a and b) each depicting only one of the two POIs: a depicts 1 (i.e., $\gamma_{a,1} \in M$, $\gamma_{a,2} \in U$) and b depicts 2 (i.e., $\gamma_{b,1} \in U$, $\gamma_{b,2} \in M$). If the player pairs a and b he/she makes a mistake, because the two photos certainly depicts different POIs. If ϵ_p is the number of errors done by a player p in a game, the player's reliability can be computed as:

$$r_p = e^{-\frac{\epsilon_p}{2}}$$

Note that r_p is a float in [0..1]: it is equal to 1 when the player makes no error, decreases to $e^{-\frac{1}{2}} = 0.6$ when the player makes 1 error, and drops almost to zero if the player makes 6 errors ($e^{-\frac{1}{6}} = 0.04$).

Considering also the player's reliability, Equations 1 and 2 respectively take the form of Equations 3 and 4:

$$s'_{b,1} = s_{b,1} + K_{pos} * r_p \tag{3}$$

$$s'_{b,1} = s_{b,1} - K_{neg} * r_p \tag{4}$$

Table 2 wraps up the decision rules that allow to increase/decrease the score of a link and to detect errors. More information about the initial value of s, and the values of K_{pos}, K_{neg}, $UPPER$ and $LOWER$ is given in Section 6.

paired with	$\gamma_{a,1} \in M$	$\gamma_{a,1} \in U$	$\gamma_{a,1} \in C$
$\gamma_{b,1} \in M$	n.a.	ϵ_p++	$s_{a,1}$++
$\gamma_{b,1} \in U$	ϵ_p++	n.a.	n.a.
$\gamma_{b,1} \in C$	$s_{b,1}$++	n.a.	n.a.

not paired with	$\gamma_{a,1} \in M$	$\gamma_{a,1} \in U$	$\gamma_{a,1} \in C$
$\gamma_{b,1} \in M$	ϵ_p++	n.a.	$s_{a,1}$- -
$\gamma_{b,1} \in U$	n.a.	n.a.	n.a.
$\gamma_{b,1} \in C$	$s_{b,1}$- -	n.a.	n.a.

Table 2. The two tables above show the decision rules used by UrbanMatch when the player pairs, or does not pair, two photos a and b on the basis of the scores of the links $\gamma_{a,1}$ and $\gamma_{b,1}$ between those two photos and a POI 1. The table on the left shows the case when the player pairs a with b, while the table on the right shows the case when the player does not pair a with b. The symbol *n.a.* means *no action*, $s_{i,j}$++ means that UrbanMatch increments the value of the score of the link $\gamma_{i,j}$ by using Equation 3, $s_{i,j}$- - means that UrbanMatch decrements the value of the score of the link $\gamma_{i,j}$ by using Equation 4, and ϵ_p++ means that UrbanMatch increases the error counter for the player p.

Functionally, the proposed solution solves the UrbanMatch problem, but an efficient solution should also consider the need for a high throughput (defined in Section 2.3). This result can be obtained by combining two approaches.

On the one hand, we can reduce the problem space by partitioning[16] the comparison space Γ. The partition is based on the concept of playable places: the comparison space Γ is built by considering only the links between each photo, retrieved in correspondence to a playable place, and all POIs visible from the same playable place. In other words, UrbanMatch discards from the comparison space Γ all the links between the photos, which were retrieved using geo-coordinates and labels of a given playable place, and the POIs in the other playable places.

On the other hand, UrbanMatch splits the set of candidates links C into two subsets: $C_{engaged}$, the set of links currently evaluated by the UrbanMatch game, whose score can be altered by players' actions, and $C_{retained}$, the set of links not yet evaluated by the game. In this way, positive and negative evidences are gathered only for links in $C_{engaged}$ whose score reaches the $UPPER$ or $LOWER$ threshold at the maximum speed. As soon as a link moves from $C_{engaged}$ to M or U, a new link is fetched from $C_{retained}$ and added to $C_{engaged}$ for evaluation.

[16] The comparison space partitioning is a well-known technique in record linkage [18].

5 The UrbanMatch Gameplay

In this section we illustrate how UrbanMatch works internally. We explain the game levels construction and the feedbacks to the players' pairing actions.

5.1 Game level definition

When the player starts the UrbanMatch app on her device, her location is detected to make her play with what surrounds her. In case of doubt (e.g. the user is close to more than one playable place), a map with the close-by locations is displayed.

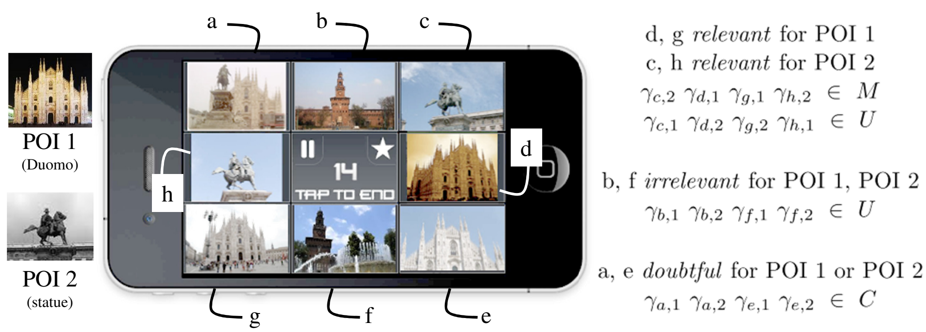

Fig. 1. Explanation of the photos presented in a game level.

Once the playable place is selected, the game starts and the players are presented with the first game level; a maximum of six levels are created and given as input to the players in each match. In each level, two POIs (1 and 2) of the playable place are considered and eight different photos (from a to h) are selected and displayed (cf. Figure 1). The photos are selected according to the following policy:

- for each of the two POIs, two *relevant photos* are definitely linked to them (d and g to POI 1, c and h to POI 2), thus representing four links belonging to the M set of correct links and four links belonging to the U set of incorrect links;
- two *irrelevant photos* are definitely linked to other POIs (b and f), not visible from this playable place; those photos are certainly not linked to the current POIs 1 and 2, thus representing four links belonging to the U set of incorrect links;
- the remaining two *doubtful photos* are not certainly linked to the current POIs 1 and 2 (a and e), thus they are representative of four candidate links from the C set of links to be validated.

The players are then asked to pair the photos depicting the same POI, but they must be careful not to select those photos referring to POIs in a different playable place, i.e. those POIs they cannot see around them.

5.2 Feedbacks to players

Whenever a player pairs two photos, this action is taken as an evidence of the correspondence between the respective links represented by those photos; each evidence is weighted according to Equations 3 and 4 and taken into consideration according to the policy defined in Table 2. Besides the application of the decision rules – which is important for the hidden purpose of data linking – the coupling action is also used to give an immediate feedback to the user within the gameplay. We decided to always give a positive or negative feedback to the player, even when UrbanMatch is in doubt; moreover, we chose to prefer a positive reward to a negative reward in case of doubt, to motivate the user to continue playing.

Whenever a user pairs two photos between the relevant and irrelevant ones, UrbanMatch always knows if the coupling action is right or wrong: either the two photos are certainly linked to the same POI – and thus the player gets a positive feedback and gains points – or they are definitely linked to different POIs – and thus the player gets a negative feedback and loses points.

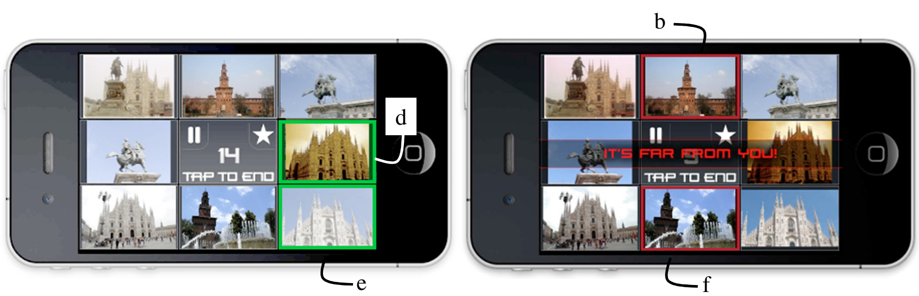

Fig. 2. Positive and negative feedbacks w.r.t. the photo pairs chosen by the player.

Every time a user pairs a doubtful photo with another one, UrbanMatch does not know if the coupling action is right or wrong, but it gives the best possible feedback: pairing a doubtful photo with a relevant one or pairing two doubtful photos gives the player a positive feedback; pairing a doubtful photo with an irrelevant one gives a negative feedback.

Figure 2 on the left shows the positive feedback given to a coupling action in a level played in "Piazza del Duomo" playable place in Milano: UrbanMatch displays a green frame around the two selected photos (d and e that both depict the Duomo cathedral of Milano) and plays a "success" sound. On the contrary, the right part of Figure 2 shows the negative feedback given to another pairing action during the same game level: UrbanMatch displays a red frame around the two selected photos (b and f, depicting Castello Sforzesco, are clearly irrelevant for this playable place since the castle is not in the playable place) together with the textual banner and plays a "failure" sound.

Each "successful" or "failure" pairing action is associated in the gameplay with a positive or negative score respectively: the sum of the scores in one level determines if the player can continue to the next level and the total score in all the levels of the match determines the position in the leaderboard.

6 Experimental Deployment and Evaluation

As reported in Section 3, in spring 2012 we experimentally deployed UrbanMatch in Milano using OpenStreetMap as POI dataset, and Flickr together with Wikimedia Commons as photo data sources. For each photo a retrieved by a query related to a POI 1, the initial score $s_{a,1}$ was set to 0.4 if the source was Flickr and 0.6 if the source was Wikimedia Commons; this was because we considered Wikimedia Commons search precision to be higher than Flickr's. For each POI i in the same playable place of POI 1, the score $s_{a,i}$ was set to 0.2, because a photo depicting a POI in a playable place may also partially show other POIs in the same playable place (e.g., see photo a in Figure 1, taken in Duomo square, that depicts both Vittorio Emanuele's statue and the Duomo cathedral). For each POI j in a different playable place, no link of the form $\gamma_{a,j}$ was inserted in the comparison space, according to the comparison space partitioning technique discussed in Section 4.

Between March and May 2012, as consequence of an email advertising campaign, seventy people downloaded UrbanMatch from iTunes and 54 of them played the game at least once, for a total of 781 played levels. The total time all players spent playing UrbanMatch is about 3 hours.

As evaluation metrics for UrbanMatch, we chose *throughput* and ALP (as defined in [31]), and *accuracy*. The latter plays an important role in deciding the values of $UPPER$ and $LOWER$. The ALP definition is equivalent to the one in [31], but the notion of *throughput* and *accuracy* need to be redefined as:

$$Throughput = \frac{CM + CU}{PlayedTime} \tag{5}$$

$$Accuracy = \frac{(CM - FP) + (CU - FN)}{CM + CU} \tag{6}$$

in which the symbols have the following meaning:

- CM is the number of candidate links that UrbanMatch was able to move from C to M, i.e., emerged as correct;
- CU is the number of candidate links that UrbanMatch was able to move from C to U, i.e., emerged as incorrect;
- FP is the number of links moved from C to M that should have been classified as incorrect, i.e., the false positive links;
- FN is the number of links moved from C to U that should have been classified as correct, i.e., the false negative links; and
- $PlayedTime$ is the total time spent by the players in playing UrbanMatch.

Note that CM and CU are a direct result of UrbanMatch, while FP and FN were manually assessed by one of the authors that lives in Milano and is thus knowledgeable about the city.

As the reader can expect, the throughput and accuracy of UrbanMatch depend on the value of $UPPER$, $LOWER$, K_{pos} and K_{neg}. Therefore we need to determine the best combination of these values to maximize both throughput and accuracy.

We arbitrarily decided that a *positive evidence* of a reliable player counts as $+0.3$ (i.e., $K_{pos} = 0.3$) and a *negative evidence* counts as -0.1 (i.e., $K_{neg} = 0.1$). We chose $K_{pos} = 3 * K_{neg}$ because players pair photos in which they recognize the same POI, but they may not pair photos for several reasons, e.g., for the little knowledge about the POI, for inexperience, for lack of time and for mistake.

To determine the best values of $UPPER$ and $LOWER$, we analysed the throughput and the accuracy of UrbanMatch as a function of $UPPER$ and $LOWER$. The values in Table 3 are obtained setting $UPPER = 1$ and assigning $LOWER$ the values 0.05, 0.10, 0.15 and 0.20[17]. Both throughput and accuracy increase when increasing the threshold, so $LOWER$ was set to 0.20.

Table 3. Throughput and Accuracy as a function of $LOWER$ threshold.

$LOWER$	0.05	0.10	0.15	**0.20**
CU	321	348	1152	**1216**
FN	4	5	7	**8**
Throughput	108.08	117.17	387.87	**409.42**
Accuracy	98.75%	98.56%	99.39%	**99.34%**

The values in Table 4 are obtained setting $LOWER = 0$ and assigning $UPPER$ values between 0.6 and 0.95 using a 0.05 step[18]. Throughput decreases while increasing $UPPER$, but accuracy increases, therefore we need to find a trade-off between the two performance indicators. Noticing that accuracy increase steeply for $UPPER \leq 0.7$ and slightly for $UPPER > 0.7$, we decided to set $UPPER = 0.7$.

Table 4. Throughput and Accuracy as a function of $UPPER$ threshold.

$UPPER$	0.60	0.65	**0.70**	0.75	0.80	0.85	0.90
CM	227	225	**68**	65	61	60	49
FP	48	47	**4**	4	3	3	1
Throughput	76.43	75.75	**22.89**	21.88	20.53	20.20	16.49
Accuracy	78.85%	79.11%	**94.11%**	95.38%	95.08%	95.00%	97.95%

[17] 0.2 is the greatest value we can assign to $LOWER$ because it is the minimum value we decided to use when initializing scores to links in C.

[18] 0.6 is the smallest value we can assign to $UPPER$ because it is the maximum value we decided to use when initializing scores to links in C.

The final results are wrapped up in Table 5. The throughput of UrbanMatch is 485 links per played hour; this is twice as much as the throughput of the ESP game [30]. The ALP of UrbanMatch is a bit more than 3 minutes per player; this is a not an outstanding result (the ALP of the ESP game is 91 minutes per player), but this value could be increased by improving the gaming and entertaining features of UrbanMatch. The accuracy of UrbanMatch is 99.06%, which is a significant result. This allow us to assert that UrbanMatch provides an effective solution for link quality assessment.

Table 5. Final evaluation results

CM	FP	CU	FN	Players	PlayedTime	Throughput	ALP	Accuracy
68	4	1216	8	54	2h 58m 12s	485 links/h	3m 17s	99.06%

7 Conclusions

The problems of interlinking and assessing the quality of information published as Linked Data have been recognized of paramount importance by researchers and practitioners, who are investigating the adoption of different approaches. Most research is focused on automated solutions, but crowdsourcing the interlinking or quality assessment tasks is also possible. Actually, if we consider the *fitness-for-use* principle of data quality [16], involving "human processors" may be the only practical way to obtain high quality links.

UrbanMatch, presented in this paper, adopts the approach of Games with a Purpose to assess the quality of automatically created links between POIs and photos that depict them. However, UrbanMatch is not simply a GWAP for Linked Data: it considers the characteristics of urban-related – or, more broadly, geo-spatially related – Linked Data and the possibility to rely on the *on-site experience* of the players in addition to their knowledge.

Our analysis of the links assessed by UrbanMatch in few month of availability on the iTunes store seems to confirm our research hypothesis. Our work and evaluation is currently oriented to gather further evidence in two directions: repeating the UrbanMatch experience in Munich[19] and exploring a different gaming approach with a new app named Urbanopoly[20].

So far, the development and deployment of UrbanMatch in the German city of Munich allowed us to verify the data preparation step, i.e. obtaining POIs from OpenStreetMap with the help of LinkedGeoData, automatically fetching photos from Flickr and Wikimedia Commons and creating the candidate links to be assessed. A preliminary analysis of the matches played in Munich confirms the results obtained in Milano.

[19] Cf. `http://bit.ly/um-munchen`.
[20] Cf. `http://bit.ly/urbanopoly`.

On the other hand, a new game named Urbanopoly was designed to get a higher value of ALP w.r.t. UrbanMatch, while keeping the same level of throughput. In analysing the results of UrbanMatch, we noticed that the players were motivated to continue playing by the presence of the leaderboard and the possibility to "beat" other players. Thus, in designing Urbanopoly, we put more emphasis on those gaming features that require a long-term engagement of the player.

The main lesson learned from UrbanMatch is that Human Computation approaches can be successfully employed to interlink urban-related datasets: the on-site experience of the players helps in gathering links with a clear business value. For example, UrbanMatch allows to learn the locations from which a POI is visible and recognizable. This information can be valuable for a wide range of city stakeholders, like municipalities (for placing information totems) or mobile operators (to deliver effective location-aware mobile advertisement). Games like UrbanMatch may serve to a wide range of Smart Cities services like traffic optimization, environmental sustainability or city planning.

Acknowledgements

This work was partially supported by Siemens Corporate Research and Technologies and by the PlanetData project, co-funded by the European Commission (FP7-257641).

References

1. G. Álvaro and J. Álvaro. A Linked Data Movie Quiz: the answers are out there, and so are the questions. `http://bit.ly/linkedmovies`, Aug. 2010.
2. S. Araújo, J. Hidders, D. Schwabe, and A. P. de Vries. SERIMI - Resource Description Similarity, RDF Instance Matching and Interlinking. *CoRR*, abs/1107.1104, 2011.
3. M. Balduini, I. Celino, E. Della Valle, D. Dell'Aglio, Y. Huang, T. Lee, S. Kim, and V. Tresp. BOTTARI: an Augmented Reality Mobile Application to deliver Personalized and Location-based Recommendations by Continuous Analysis of Social Media Streams. *Journal of Web Semantics, accepted with minor revision*, 2012.
4. C. Batini, C. Cappiello, C. Francalanci, and A. Maurino. Methodologies for data quality assessment and improvement. *ACM Comput. Surv.*, 41(3), 2009.
5. O. Benjelloun, H. Garcia-Molina, D. Menestrina, Q. Su, S. E. Whang, and J. Widom. Swoosh: a generic approach to entity resolution. *VLDB J.*, 18(1):255–276, 2009.
6. A. Caragliu, C. Del Bo, and P. Nijkamp. Smart cities in Europe. Serie Research Memoranda 0048, VU University Amsterdam, Faculty of Economics, Business Administration and Econometrics, 2009.
7. I. Celino, S. Contessa, M. Corubolo, D. Dell'Aglio, E. Della Valle, S. Fumeo, and T. Krüger. UrbanMatch – linking and improving Smart Cities Data. In *Linked Data on the Web Workshop (LDOW2012)*, 2012.
8. I. F. Cruz, C. Stroe, F. Caimi, A. Fabiani, C. Pesquita, F. M. Couto, and M. Palmonari. Using AgreementMaker to align ontologies for OAEI 2011. In Shvaiko et al. [24].
9. E. Della Valle, I. Celino, and D. Dell'Aglio. The Experience of Realizing a Semantic Web Urban Computing Application. *Transactions in GIS*, 14(2), 2010.

10. E. Della Valle, I. Celino, D. Dell'Aglio, F. Steinke, R. Grothmann, and V. Tresp. Semantic Traffic-Aware Routing for the City of Milano using the LarKC Platform. *IEEE Internet Computing*, 15(6), 2011.
11. J. Euzenat, A. Ferrara, W. R. van Hage, L. Hollink, C. Meilicke, A. Nikolov, D. Ritze, F. Scharffe, P. Shvaiko, H. Stuckenschmidt, O. Sváb-Zamazal, and C. T. dos Santos. Results of the ontology alignment evaluation initiative 2011. In Shvaiko et al. [24].
12. J. Euzenat and P. Shvaiko. *Ontology matching*. Springer-Verlag, Heidelberg (DE), 2007.
13. I. P. Fellegi and A. B. Sunter. A Theory for Record Linkage. *Journal of the American Statistical Association*, 64(328):1183–1210, 1969.
14. A. Flemming and O. Hartig. Quality Criteria for Linked Data sources. Online at `http://bit.ly/ld-quality`.
15. T. Heath and C. Bizer. *Linked Data: Evolving the Web into a Global Data Space*. Synthesis Lectures on the Semantic Web. Morgan & Claypool, 2011.
16. S. A. Knight and J. Burn. Developing a framework for assessing information quality on the World Wide Web. *Informing Science*, 8, 2005.
17. T. Lee, S. Park, Z. Huang, and E. Della Valle. Toward Seoul Road Sign Management on the LarKC Platform. In *Posters and Demos of ISWC 2010*, 2010.
18. A. McCallum and B. Wellner. Object consolidation by graph partitioning with a conditionally-trained distance metric. *Proceedings of the ACM Workshop on Data Cleaning, Record Linkage and Object Identification*, 2003.
19. F. Naumann and C. Rolker. Assessment Methods for Information Quality Criteria. In B. D. Klein and D. F. Rossin, editors, *IQ*. MIT, 2000.
20. H. B. Newcombe. *Handbook of record linkage: methods for health and statistical studies, administration, and business*. Oxford University Press, Inc., 1988.
21. X. Niu, S. Rong, Y. Zhang, and H. Wang. Zhishi.links results for OAEI 2011. In Shvaiko et al. [24].
22. E. Rahm and P. A. Bernstein. A survey of approaches to automatic schema matching. *VLDB J.*, 10(4):334–350, 2001.
23. T. C. Redman. *Data Quality: The Field Guide*. Digital Press, 2001.
24. P. Shvaiko, J. Euzenat, T. Heath, C. Quix, M. Mao, and I. F. Cruz, editors. *Proceedings of the 6th International Workshop on Ontology Matching*, volume 814. CEUR-WS.org, 2011.
25. K. Siorpaes and M. Hepp. Games with a Purpose for the Semantic Web. *Intelligent Systems*, 23:50–60, 2008.
26. C. Stadler, J. Lehmann, K. Höffner, and S. Auer. LinkedGeoData: A Core for a Web of Spatial Open Data. *Semantic Web Journal*, 2011.
27. J. Volz, C. Bizer, M. Gaedke, and G. Kobilarov. Discovering and Maintaining Links on the Web of Data. In *International Semantic Web Conference*, volume 5823 of *LNCS*, pages 650–665. Springer, 2009.
28. L. von Ahn. Games with a Purpose. *IEEE Computer*, 39(6):92–94, 2006.
29. L. von Ahn. Human Computation. In G. Alonso, J. A. Blakeley, and A. L. P. Chen, editors, *ICDE*, pages 1–2. IEEE, 2008.
30. L. von Ahn and L. Dabbish. Labeling images with a computer game. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, CHI '04, pages 319–326, New York, NY, USA, 2004. ACM.
31. L. von Ahn and L. Dabbish. Designing games with a purpose. *Commun. ACM*, 51(8):58–67, 2008.
32. W. E. Winkler. Overview of record linkage and current research directions. Technical report, Bureau of the Census, 2006.