Leveraging Web 2.0 for Informed Real-Estate Services

Papantoniou Katerina*, Athanasiadis Marios - Lazaros*, Fundulaki Irini*, Georgis Christos*
Stavrakas Yannis**, Troullinos Michalis**, Tsitsanis Anastasios***

*Institute of Computer Science, FORTH-ICS, Greece
{papanton,mathanas,fundul,georgis}@ics.forth.gr,
**Institute for the Management of Information Systems, Research Center - ATHENA
{yannis,mtroullinos}@imis.athena-innovation.gr
***TREK Consulting
t.tsitsanis@trek.gr

Abstract. The perception about real estate properties, both for individuals and agents, is not formed exclusively by their intrinsic characteristics, such as surface and age, but also from property externalities, such as pollution, traffic congestion, criminality rates, proximity to playgrounds, schools and stimulating social interactions that are equally important. In this paper, we present the Real-Estate 2.0 System that in contrary to existing Real-Estate e-services and applications, takes also into account important externalities. By leveraging Web 2.0 (content from Social Networks, POI listings) applications and Open Data enables the thorough analysis of the current physical and social context of the property, the context-based objective valuation of RE properties, along with an advanced property search and selection experience that unveils otherwise "hidden" property features and significantly reduces user effort and time spent in their RE quest. The system encompasses the above to provide services which assist individuals and agents in making more informed and sound RE decisions.

1 Introduction

Contemporary real estate economists (Geltner and H., 2007),(Van Dijk et al., 2011) agree that the value of an apartment, building or land in an urban area is not only represented exclusively by the intrinsic characteristics of the property, such as the quality, location and size of its construction, but also by both positive externalities such as proximity to playgrounds, schools, dynamic local cultural relations, intellectual circuits of exchange, peaceful and stimulating social interactions; and negative externalities such as air and noise pollution, traffic congestion, noisy neighbours, high criminality rates.

Real-Estate 2.0, in contrary to existing Real Estate web services and applications (xe [1], foxtons [2]) that are restricted only to intrinsic characteristics of properties, aims to harvest and integrate various forms of Real Estate (RE) information that is actually available today in various Web 2.0 sources, under the form of point of interest (POI), content and opinions expressed in

1. http://www.xe.gr/
2. http://www.foxtons.co.uk/

Social Web and other relevant information published as Open Data. The goal of the system is threefold, first to provide highly personalized access through the integrated RE information space by means of rich spatial, temporal and content related constraints, second to support multi-criteria analysis and evaluation of all site location aspects so as to affect final strategic investment decisions and third to provide Real Estate analytics and reporting, taking advantage of the character of available data.

The first objective towards this direction is related to the acquisition of four distinct kinds of RE data: (a) property features, (b) POIs in the surrounding area of properties, (c) social content under the form of discussion posts from online Social Media, that provide material evidence regarding the physical and social context of the properties and (d) Open Data mostly from government sources that provide official statistics, demographics and reports about various aspects of life in an area.

Real-Estate 2.0 relies on properties data available by real estate agents through their in-house databases. Those data are complemented with additional property data by crawling main RE advertisement sites and portals. Regarding the acquisition of POIs, Real-Estate 2.0 harvests the relevant data from a multitude of administrative and touristic web sites including freely available databases of GPS data, or social content related to POIs. The main challenge in this respect is the ability to steer the acquisition process to POIs in or close to a certain geographical area of interest, as well as the detection of new POIs as time evolves.

Social content may be acquired from several sources, such as blogs, forums, or social platforms such as Google Places and Foursquare. While blogs and forums are normally accessed as Web pages, the access to Google Places and Foursquare is different as these sources give limited access to their own data through their APIs.

Open Data is about knowledge that can be used, reused and redistributed. The last years have seen an enormous effort in publishing open data from scientific and government sources. Initiatives such as Data.gov.uk [3], NYC Open Data [4] and OECD [5] provide a plethora of datasets, usually in different formats, so the acquisition and the integration in our system was another challenging task.

The integration of the acquired RE information is actually the second objective and will determine the quality of the data actually exploited by Real-Estate 2.0 services. First, property information needs to be integrated with POIs. The subtle issue in this respect is related to the precision of the geographical information available in the ads (e.g. associated with neighbourhood or broader areas instead of specific addresses). Such a situation is common when property ads are obtained automatically from sources by means of focused crawling.

Another challenge is to associate POIs with social content from social networks. Users do not necessarily refer to POIs in a consistent way, but may use abbreviations, colloquial language, which is harder to analyze automatically. Also, content generated by users may be short (for example, up to 140 characters in the case of Twitter), or may link to Web pages containing additional information or anonymized opinions.

An additional objective relates to the implementation of end user services towards the realization of the goals previously mentioned. The system must enable users to explore the integrated and longitudinal information by searching for properties satisfying a set of given

3. http://data.gov.uk/
4. https://nycopendata.socrata.com/
5. http://www.oecd.org/

spatiotemporal constraints and individual user preferences concerning the quality of life and socio-economic needs.

To summarize, Real-Estate 2.0 system provides:

— a generic infrastructure for harvesting, extracting, aggregating and curating contextual information related to RE properties from administrative, touristic or social media sites (e.g., discussion forums, social networking sites), which allows us to reduce the cost of acquiring and maintaining high quality location data over time (Batini et al., 2009);
— an interface for agents and individuals to explore spatio-temporal trajectories of the areas surrounding the properties or points of interest, in order to assess positive and negative externalities of properties;
— advanced business models for real-estate agencies to manage their portfolio based on past property value evolution (Anselin, 1988) and location-related indices (Holly et al., 2010a),(Holly et al., 2010b)

In the remainder of this paper, we exemplify the vision of the system through use case scanarios while in the section 3 we describe the architecture of the system and we analyze each layer. Then we mention some related work and conclude with a discussion of the contributions and future work.

2 Case Studies

In the following, we exemplify the main vision of the system by two core usage scenarios highlighting how Real-Estate 2.0 system could potentially (a) assist individuals in buying home residences and (b) manage a portfolio of commercial and residential properties or lands as part of corporate RE investment optimization.

Comparative RE Market Analysis: Imagine Maria, a female white-collar worker who needs to move to a new city because of her job. Being in a city that is completely unknown to someone, and wanting to buy (or even rent) a family home can be an extremely difficult task. Maria has already some requirements that emerge from family, work, as well as personal preferences. She needs to be close (or within walking distance) from a metro station so that she can take the metro to work every day. On the other hand, her husband needs to drive to work so he needs to have either a parking space in the building or access to parking in the street. Additionally her two kids, a boy and a girl aged 10 and 15 respectively, need to be close to school and have access to nearby facilities for their extra curriculum activities like swimming, foreign languages lessons, ballet, etc.

Having to actually aggregate all these personal or group preferences is obviously a tedious task; in order to be able to fulfill all the requirements the required information should not be restricted to availability only but should also include a discussion on quality: there are ballet schools but how good are they, there are parking places on the street but how safe is the neighbourhood for street parking, there is a school within walking distance but how good is it, there is a super-market but what is the quality of the products, there is a playground in the area but is it safe / large enough? Moreover one may need to find information about things that are specific in the area: having a football field next to one's house might be excellent for the children for exercise but could also be a big trouble if it is also used by a football team that has fans that occasionally create trouble around football games.

Thus we can easily recognize that Maria will benefit from a service that utilizes a more broader

pool of data (POI, social content and Open Data) apart from simple listings of available RE properties in a region. Location is not the only property of RE assets, the social and physical environment and its evolution over time play also an important role. The integrated information makes it possible to analyze the evolution of real-estate values.

Intelligent RE Portfolio Management and Optimization: Imagine House - Investments SA, an international real estate investment company, with a portfolio of over 1.000 commercial, residential properties, and land items in Greece, located across diversified markets (regions and cities) of the country in order to facilitate management and mitigation of potential investment risks. The company's management board has decided and set specific objectives for the next 10 years, concerning the performance of the property and land holdings and the return on the investments that have made (and will make) towards building the company's real estate portfolio. Towards the realization of these objectives, the company has developed a balanced strategy that includes the following actions:

- Selling directly properties and land items in regions with important land and property values reduction projections over the next 10 years.
- Re-investing in new properties and land items purchases in regions with projected value increase over the next 10 years and selling them back at the projected value peak period for maximizing profits.
- Leasing properties and land items in regions with a significant expected rise in leasing.
- In addition, leasing will be applied in all these properties of the 1st category that will not have been sold within the first two years of the strategy implementation.

In each of the aforementioned cases, the company needs to assess projections concerning property, land and leasing values in different regions of the Greek territory for optimizing the portfolio performance and maximizing the upcoming investments' returns. These projections will be based on the analysis and further processing of historical geo-referenced data regarding the aforementioned values (how they have performed in past periods), and through their correlation with cyclic indicators such as capital appreciation and GDP growth for locations and properties chosen and portfolio weights (the characteristics that affected the value changes), sound estimations will be made on the projected future property values for different time periods.

3 Real-Estate 2.0 System Architecture

The overall architecture of the system is depicted in figure 1. There are three layers, namely RE services, integration and crawling. RE services layer encompasses modules that implement novel real-estate e-services using state of the art Web 2.0 technologies as well as the integrated data and access services supported by the Real-Estate 2.0 system. The integration layer is responsible for the consolidation, georeferencing and management of the harvested data in a way that enables spatio-temporal and uniform access to current and archived content. Last, the crawling layer makes available to system data though systematic and focused crawling of the Web. The communication between layers is achieved through APIs while the overall system follows a service-oriented logic so as to function as a stand-alone system and with the provision of integration with existing systems.

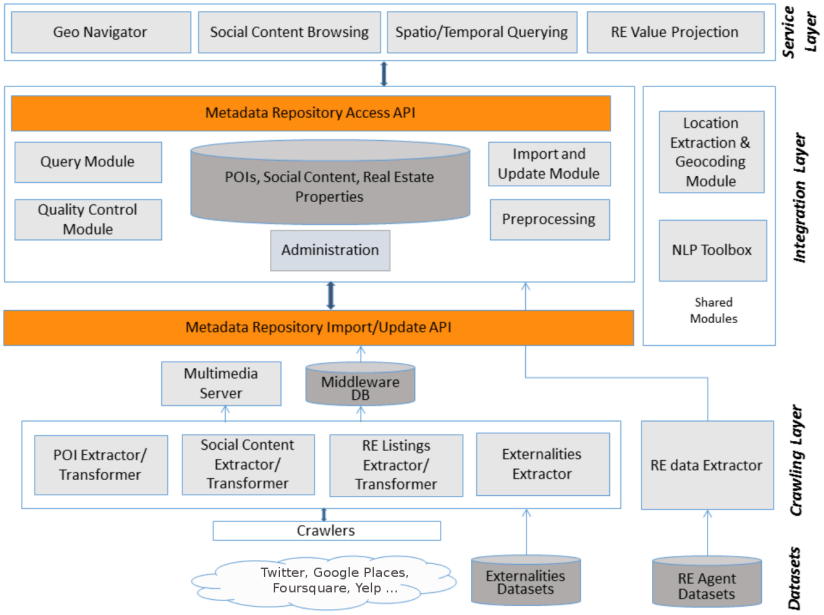

FIG. 1: Real-Estate 2.0 Architecture.

3.1 Crawling Layer

The crawling layer is responsible for launching and coordinating focused crawls in order to extract, store, and make easily accessible all the information that constitute the externalities of the real estate items within a specific area. The crawling subsystem targets three distinct types of information: POIs, social comments, and Open Data. For each information type, a number of sources that provide this information are supported. For example, POIs are collected by querying the APIs of global services like Foursquare and Google Places. Our prototype implementation aims to cover two metropolitan areas, London and Athens, therefore local POI sites are also queried during the data collection process. The retrieval process is limited to only those POIs that are within a specified geographical area. Social comments come from two categories of sources: POI sources (mentioned before), and social networks like Twitter. Again, the challenge is to link relevant comments with the area they refer to. In the case of POI sources this is straightforward, since user comments are directly associated with a specific POI. By contrast, in the case of tweets special language analysis tools must be used in order to establish a relationship between a tweet and a geographical area or a POI. Finally, Open Data retrieved by the crawling subsystem are usually demographic or statistical data provided by state authorities that provide information about areas like pollution, criminality, or traffic congestion. The extensive heterogeneity of the Open Data formats and retrieval methods require specialized access method for each type of dataset. Therefore one cannot implement a solution as general as for POIs and social comments, but go for an approach tailored to a specific area and specific data sources. The crawling subsystem focuses on the three types of information

described above. The pivotal concept is that of a data collection campaign. A campaign is set by defining a number of parameters, which can be accomplished either programmatically through an API, or interactively through a user interface (called the Crawler Cockpit). Those parameters describe, first and foremost, the geographical area of interest in the form of a circle (center and radius). This circle is converted to whatever the data sources understand, in order to minimize the retrieval of irrelevant information. Other campaign parameters allow for: (a) The selection of POI categories to be included in the campaign. POIs are classified in a set of predefined categories, like for example schools, parks, restaurants, etc. (b) The definition of the time interval for the crawling of each POI category. Each POI category is associated with a time interval that defines how often the campaign must refresh the particular POIs by re-launching the respective crawlers. (c) The selection of the media types for each retrieved item, including text, image, and video. Our solution faced a number of technological challenges, including: (a) support for concurrent campaigns lasting up to months, (b) respecting source constraints on the number of allowed requests per time unit, and (c) coordinating a large number of active crawler instances. Campaign results are stored in a repository consisting of a relational database for structured data and metadata, and a directory structure for keeping the media items. The Crawler Layer API offers a flexible way to retrieve campaign data, by setting a number of criteria that the results must satisfy. Since a campaign may run for long periods of time, an incremental method for retrieving only the latest information is also available to the Integration Layer.

3.2 Integration Layer

The integration layer provides a syndicated view of information (POIs, Open Data, social content and real-estate properties) stored in a centralized database allowing the consolidation of all information based on spatial and temporal predicates.

The Real-Estate 2.0 integration layer schema is based on the W3C Points of Interest Data Model (Hill and M., 2012), the W3C recommendation for the modelling of POIs on the World Wide Web. In the context of this model, a POI is defined as a set of loosely coupled and inter-related geographical terms, comprised of Locations, POIs and Places. Based on the above terminology, the POI data model consists of a `POI` entity that has a number of properties for capturing descriptive information (e.g. category, external links, description, price ranges, opening and closing hours) along with a `Location` entity describing its location. The Real-Estate 2.0 schema follows in general terms W3C POI Data Model and considers the entities `POI` and `Location` as the integration hub for all the available data. More specifically, a real estate property - a human construct with fixed location, open to the market for rent or sell - is conceived as a special type of a POI with different characteristics. The W3C POI Data Model assumes no distinction between a `Place` and a `POI` because they share the same attributes although often with differing interpretations based on scale. In analogy, we also consider `REProperty` as a special type of `POI` with small differences in attributes interpretations (e.g. different categorization). This enables the uniform integration of externalities (e.g. social content, Open Data) with the core entities of the model (POI, REProperty). For the representation of geospatial information we adopted the WGS84 (World Geodetic System) standard that is defined by a longitude, latitude and elevation triple. This geodetic system allows complex geospatial queries (e.g. proximity of real estate properties to POIs, the value of criminality in an user-defined polygon) and provides flexibility in the representation of entities beyond unique points through

polygons and lines (e.g. a polygon for a national park, a line for an arterial road) in comparison with other representation for example postal addresses. The elevation was eliminated in our model because of the absence of altitude information in available data.

For the modelling of social content and Open Data we adopted a uniform approach with the introduction of the `Externality` entity despite the high heterogeneity in sources and format of the available data ranging from free text (as in comments, posts etc.), to multimedia content, ratings (e.g. likes in Facebook, check-ins) and measurements in different metric systems. An externality instance can be linked either to `POI` or a `Real Estate Property` entities. This requires georeferencing for the linking of the externalities to a location entry and normalization of different data values. Last but not least, the representation of provenance is particularly important given the dependence of the system with looser sources such as social commentary on the Web. In traditional databases and generally to electronic documents the reliability and integrity of data considered as a given, a case that cannot be taken for granted in the Web. The consumer of knowledge is of paramount importance to be aware of the origin of data. For this purpose, the entity `Provenance` maintains the required attributes (author, source, first publication date. license etc.). Based on the general proposed principles of modeling provenance data in our model we have adopted the approach in which data and annotations coexist in the same system, adding additional information to database but without the need of recalculation (Cheney et al., 2009). A depiction of the core entities and their relationships is given in figure 2.

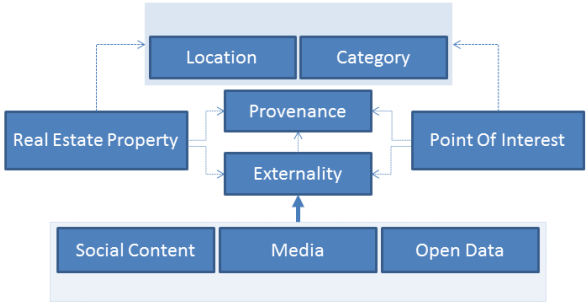

FIG. 2: High Level Real-Estate 2.0 Schema.

After the schema integration follows the loading process with data from the sources. This is a complex and expensive process that includes customization integration of the heterogeneous sources into a common format, cleaning and quality control of the resulting dataset according to specific business rules and finally the loading to the repository. For all these reasons, we adopted an approach in which data transformations take place outside the database server. The strengths of this approach is the reduction in development time as only data relevant to the solution is extracted and processed by the database server, while the database server remains free from preprocessing tasks until a bulk new update occurs. As far as update mechanism is concerned, the integration layer can schedule campaigns through the crawling layer in order to aggregate data asynchronously. Updating tasks are employed off-line when the system passes to a maintenance mode. Update policy differs according to the type of updated data for example

real estate property updates and archived tasks are scheduled on a more frequent basis than Open Data. Subsequently, we describe the modules for quality assurance and text processing tasks.

Quality module

Data quality is not an option but a constraint for the proper operation of the repository. Data quality problems seem to introduce even more complexity and computational burden to the loading process of the repository. This module is responsible for treatment of data duplications and noise in extracted POIs through heuristics and similarity measures. The process starts with conflict resolution based on source reliability (e.g. Google Maps is considered the most reliable source). In the second stage, problems with missing categories are resolved using context information while in the third stage longitude and latitude values are cross-checked with the address of the POI through geocoding and reverse geocoding. Finally, a weighted combination of string similarity measures such as Leveinstein, NeddlemanWunch, SmithGooth, Soundex to name a few, leads to the appropriate CRUD action.

Text Processing

A major part of the available data, that contain invaluable information, is provided in an unstructured plain text format such as real estate property descriptions and social content. From a real estate property description, information about characteristics and amenities can be extracted such as number of bedrooms, number of bathrooms, region that the property resides (e.g. address, neighbourhood), heating / cooling facilities etc. In addition, from the innumerable production of social content those referring to a region, neighbourhood or a POI must be detected so as aspects and opinions about the specific location to be revealed. To accomplish that the raw text is analyzed through a pipeline of natural language processing tools. The pipeline includes language identification, sentence segmentation, text normalization, part of speech tagging, extraction of candidate named entities through patterns and finally a disambiguation process. The disambiguation process is necessary as a precision improvement step in order to eliminate false positive annotations and for the resolution in cases where multiple locations share the same name. For the disambiguation step major Web knowledge databases (DBPedia, Geonames) along with context and profile information (if existent) were used. This process was very challenging due to the short, informal and usually allusive nature of text. On top of that, we handled text written in two languages namely English and Greek. The degree of difficulty is increased in case of Greek because obviously is a less studied language from a computational perspective with limited number of available tools.

3.3 Services Layer

The Services Layer accommodates four core modules that enable the provision of first-of-a-kind added value RE end-user services, on the basis of proven Web 2.0 technologies and techniques. In more detail the Geo-Navigator module enables the integration with third-party services (e.g. Google Maps, Street View, OpenStreetMaps), allowing remote exploration of RE properties in interest and visual representation of social and spatio-temporal content around them. Such content is fed in the Real-Estate 2.0 system through the Social Content Browsing module, providing direct access to comments and opinions expressed in popular social media regarding property-surrounding areas and nearby POIs, along with the Spatial / Temporal Querying Module that handles the formulation of spatial and temporal predicates involving POIs, Open Government Data and property features. The latter represents the main input

source for the RE Value Projection Module which exploits a wide variety of datasets (native property characteristics and relevant spatial features, location-based contextual and social data, spatio-temporally characterized open data) to enable and mobilize a multiple regression analysis process, which in turn defines the Real-Estate 2.0 Hedonic Pricing Model for the objective valuation and short term projection of selected RE properties' values. Based on core services more sophisticated services will build in order to provide more advanced functionality as in detail described in use case scenarios.

3.4 Implementation

From implementation perspective, for the repository of the integration layer the PostgreSQL DBMS was selected in combination with PostGIS extension to support the representation of geographic objects and the subsequent implementation of complex spatial queries. In addition, the ltree extension of PostgreSQL was used for the representation of the hierarchy of POI categories for the optimization of related queries. For the georeferencing tasks (e.g geocoding and reverse geocoding) we rely on third parties applications namely Google Geocoding [6] and the collaborative Nominatim [7]. The communication between the layers achieved through RESTful web services ensuring the interoperability and extensibility of the system. The major part of the system is built with Java technologies.

3.5 System Demonstration

In the following section, we present briefly some of the functionality of the first prototype of our system [8]. The first screenshot depicts the multi-parameter real estate properties search options (region, price criteria, amenities etc.) and an overview of search results. The interface enables the user to apply proximity and complex searches over the consolidated information space. At the second screenshot, the enriched search results are depicted: (a) an interactive

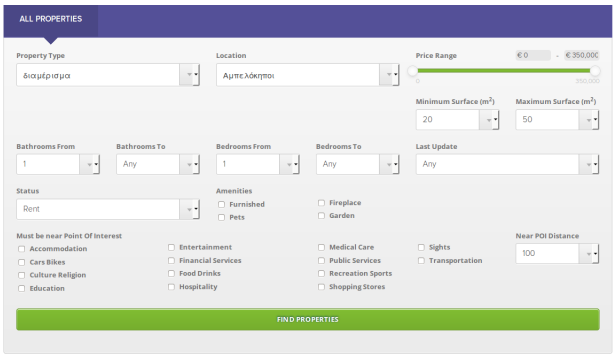

FIG. 3: Real-Estate 2.0 Search options

6. https://developers.google.com/maps/documentation/geocoding/
7. http://wiki.openstreetmap.org/wiki/Nominatim/
8. http://re2-vm-win.imis.athena-innovation.gr/

map where approximate POIs to the real estate property are depicted according with the user preferences in the right menu with (b) nearby POI categories (c) RE property multimedia (d) related Open Data (e) overview (f) detailed description (g) amenities.

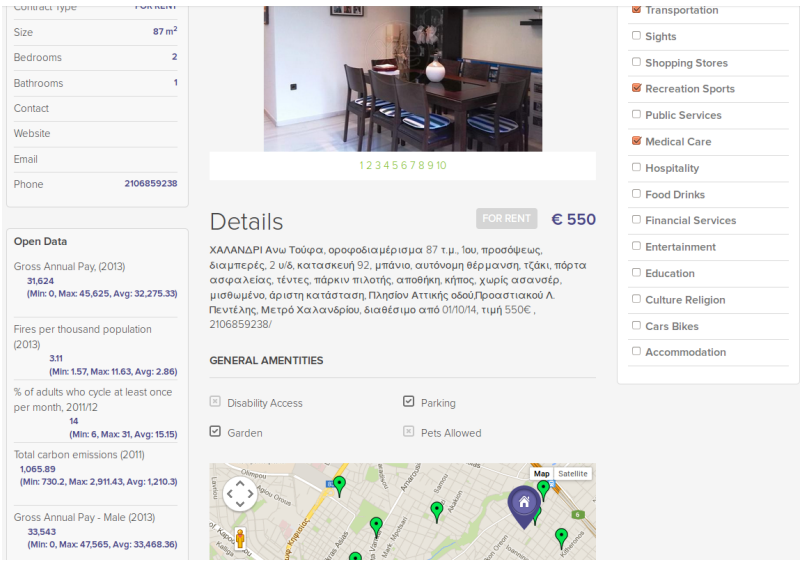

FIG. 4: Real-Estate 2.0 Results Overview

4 Related Work

Even though the real estate market is one of the major moving forces worldwide, IT penetration in RE agents services portfolio is still poor. The current status in small (mainly local) real estate agencies is the web presence through simple websites showing static data and pictures about their property listings and with search functionalities based on simple, property-specific criteria, lacking map visualizations or POIs exploration. Larger RE agents and banks, have developed and use internal software applications that process limited, static and rather complex data volumes referring mainly to statistics concerning RE financial activity, transaction rates, value zones and legal data. The extroversion of these services is limited to the provision of statistical studies concerning the evolution and financial aspects of the real estate market. Finally, the major property listing sites offer rather limited search services, which do not cover the full spectrum of customers' needs for information provision, and thus, not allowing for and supporting the establishment of an integrated framework for added value RE services provision that minimizes customer effort and facilitates RE transactions and investments. So, to the best of our knowledge, this is the first system which integrates in real estate business models both traditional real estate data with user-generated content from online social networks, modelling the externalities that affect the value of real estate properties. The specific mechanisms we develop in Real-Estate 2.0 are spanning in many different areas

including: social comment extraction, POI recognition and enrichment, entity matching, natural language processing, spatiotemporal data management, Web crawling and Web archiving, house price index estimation.

In the context of real estate price estimation a recent work is that of (Sun et al., 2014) that proposes a new method of real estate price index prediction by introducing human behavioural factor into the forecasting model. For this purpose, they combine sentiment from online daily news and Google search engine query data, so as to construct an integrated data mining model that has forecasting abilities.

Next, in the area of POI recognition, enrichment and social content extraction, a related work is that of (Cano et al., 2011) that demonstrates a methodology for modelling the collective perception of a POI. In this work, after the phases of the social content extraction and the subsequent semantic enrichment of POIs by exploiting Web 2.0 applications, they perform a triplification of POIs through their LinkedPOI ontology to enable a visual representation of POIs.

There is a large volume of work in recent years for the extraction of localized information from social content text. Many of them are based on Natural Language Processing but their works are only applicable in English language. One exception, is the work for the Portuguese language of (Santos et al., 2012) that presents multiple ways of places enrichment, applying Natural Language Processing and Information Extraction techniques on contents fetched from the web and additionally exploits Web knowledge bases such as Wikipedia and Wiktionary as a validation mechanism.

Finally, for the POI deduplication and entity matching, there are plenty of proposed methodologies, one of the most recent that demonstrates high precision and recall results is the (Zheng et al., 2010) that based on a machine learning based approach. In this approach, the proposed features in order to find the differences between two POIs, were name similarity, address similarity, category similarity, as well as corresponding metrics. In our system we created some scenarios based on POIs categories in order to avoid expensive machine learning techniques.

5 Conclusions and Future Work

This paper has presented Real-Estate 2.0 approach from a system's design perspective and focused on our solution for acquisition, integration and the querying of real-estate related data towards the implementation of novel and advanced RE e-services such as exploration of RE properties using spatial, temporal and content related preferences, property analysis using location context and real-estate analytics over time. In our context real estate data are not restricted to intrinsic characteristics of a property but include important aspects such as POIs, social content and Open Data. The overall approach leverages Web 2.0 in multiple ways such as for data acquisition, georeferencing, visualizations etc. The ongoing work encompasses an extensive user based evaluation as well as improvements for individual methods (e.g. quality module, house price projection) and implementation of more advanced end-user services. From a socioeconomic point of view, the system aims to provide a competitive edge to real estate agencies by offering custom geo-data services that can incorporate to existing infrastructure with a minimum effort cost and without any up-front capital or maintenance investment.

6 Acknowledgement

This work was supported by the national "COOPERATION 2011" programme, project with code 11SYN_1_531 entitled "Informed Real-Estate Services: Leveraging Web 2.0".

References

Anselin (1988). *Spatial Econometrics: Methods and Models*. Boston: Kluwer Academic Publishers.

Batini, C., C. Cappiello, C. Francalanci, and A. Maurino (2009). Methodologies for data quality assessment and improvement. *ACM Comput. Surv. 41*(3), 16:1–16:52.

Cano, A. E., G. Burel, A.-S. Dadzie, and F. Ciravegna (2011). Topica: A tool for visualising emerging semantics of pois based on social awareness streams. *10th International Semantic Web Conference*.

Cheney, J., L. Chiticariu, and W.-C. Tan (2009). Provenance in databases: Why, how, and where. *Found. Trends databases 1*(4), 379–474.

Geltner, D. and P. H. (2007). A set of indexes for trading commercial real estate based on the real capital analytics transaction prices database, mit center for real estate. Technical report, MIT Center for Real Estate, Commercial Real Estate Data Laboratory - CREDL.

Hill, A. and W. M. (2012). Points of interest core. Technical report, W3C.

Holly, S., M. Pesaran, and T. Yamagata (2010a). A spatio-temporal model of house prices in the usa. *Journal of Econometrics 158*(1), 160–173. M1 - 1.

Holly, S., M. H. Pesaran, and T. Yamagata (2010b). Spatial and Temporal Diffusion of House Prices in the UK. Technical report.

Santos, J., A. Alves, F. C. Pereira, and P. Abreu (2012). Semantic enrichment of places for the portuguese language. *INForum2012*.

Sun, D., C. Zhang, W. Xu, M. Zuo, J. Zhou, and Y. Du (2014). Does web news media have opinions- evidence from real estate market prediction. *18th Pacific Asia Conference on Information Systems PACIS 2014*.

Van Dijk, B., P. Franses, R. Paap, and D. van Dijk (2011). Modeling regional house prices. *Journal of Applied Economics 43*, 2097–2110.

Zheng, Y., X. Fen, X. Xie, S. Peng, and J. Fu (2010). Detecting nearly duplicated records in location datasets. *18th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*.