A knowledge discovery mechanism to user requirement identification in building design

Identificación de los requisitos del usuario en el sector de la construcción bajo mecanismos de descubrimiento del conocimiento

Fernanda Flach*, Marco Aurélio Stumpf González ^{1**}, Andrea Parisi Kern***

* Universidade do Vale do Rio dos Sinos (UNISINOS), São Leopoldo, RS. BRASIL

Fecha de Recepción:16/04/2012

Fecha de Aceptación:20/07/2012

PAG 83 - 98

Abstract

The purpose of this paper is to investigate how the knowledge of real estate market can be used to support user requirement identification. A construction project well adjusted to the user requirements increase value and causes minors changes during its life cycle. As a consequence, renewal, refurbishments, and demolition are less present, reducing waste generation, reworking and material consumption. It is especially important in housing customization markets. However, one of the challenges faced by designers is frequently concerned about how properly to identify user requirements, wishes and needs, which are on the essence of the briefing phase. In this context, real estate data can be useful to designers, since it reflects the users' evaluation of the building attributes. The research strategy uses a knowledge discovery mechanism, composed by five steps: (1) formulation of a general database; (2) specific data selection using Case-Based Reasoning; (3) enrichment of data-sample; (4) development of hedonic price models using regression analysis; and (5) simulation of the value of design alternatives. Based on an application of an hedonic price model, using data from the medium-class housing market of Porto Alegre, Brazil, the main results indicate that adjusted price models have sufficient detailing and statistical precision to support decisions in the initial stage of design.

Keywords: User requirements, housing customization, design process, knowledge processes, hedonic price modeling

Resumen

El propósito de este trabajo es investigar la forma en que el conocimiento del mercado inmobiliario puede ser utilizado para ayudar a identificar los requisitos de los usuarios. Un proyecto de construcción más ajustado a las necesidades de los usuarios puede aumentar su valor y producir cambios menores durante su ciclo de vida. En consecuencia, la renovación, remodelación y demolición estarán menos presentes, reduciendo la generación de residuos, la re-construcción y el consumo de materiales. Esto es especialmente importante en los casos de personalización de la vivienda. Sin embargo, uno de los desafíos que enfrentan los diseñadores es cómo identificar adecuadamente las necesidades de los usuarios, sus deseos y necesidades, pues constituyen un factor fundamental en el proceso inicial de información. En este contexto, los modelos estadísticos de datos basados en los bienes raíces pueden ser de gran utilidad para los diseñadores, pues reflejan la evaluación que poseen los usuarios de los atributos de construcción. La estrategia de investigación usa un mecanismo de descubrimiento del conocimiento, compuesta por cinco pasos: (1) la formulación de una base de datos general, (2) la selección de datos específicos utilizando el Razonamiento Basado en Casos, (3) el enriquecimiento de los datos, (4) el desarrollo de modelos de precios hedónicos utilizando el análisis de regresión, y (5) la simulación del valor de las alternativas de diseño. Sobre la base de aplicación de un modelo de precios hedónicos y utilizando los datos del mercado de las viviendas de clase media de Porto Alegre, Brasil, los principales resultados indican que los modelos de precios ajustados contienen suficientes detalles y precisión estadística para apoyar las decisiones en la etapa inicial de diseño

Palabras Clave: Requisitos de los usuarios, personalización de las viviendas, proceso de diseño, proceso de conocimiento, modelos de precios hedónicos

1. Introduction

The segment that involves the activities of architecture, engineering, and construction (AEC industry) has great importance for the society, regarding economic, social, and environmental aspects. A construction project well adjusted to user requirements increases building value, and probably it will have minor changes during its life cycle (Koskela, 2000). However, to identify user requirements is not a simple task for designers. This is especially important in housing customization markets, such as pre-sales segments (Juan et al., 2006; Shin et al., 2008). The challenge faced by designers is concerned about how to identify user requirements, wishes, and needs, since the initial stage of building design, known as briefing (RIBA, 2007).

¹ Autor de correspondencia / Corresponding author:
E-mail: mgonzalez@unisinis.br

The purpose of this paper is to investigate how the knowledge of real estate market can be used to support user requirement identification. We present a knowledge discovery mechanism designed to identify user's requirements and to support decision-making process. The proposal is based in hedonic price models which are generated through market transaction data. Paper discusses briefly design, actual systems to support briefing, and related aspects. In sequence, it is presented the proposal and some results obtained from a case study, concerning price modeling of middle-class market range apartments, using data from Porto Alegre, a southern Brazilian city.

2. Building design process

2.1 The role of user' requirements and the initial steps of design

Building design also can be regarded as a complex task. By its nature, building design is a creative process where problems and solutions emerge simultaneously, in rhetoric, persuasive, and exploratory ways. It requires the identification and weighting of different needs, requirements and wishes, which need be properly translated to the construction language to be incorporated into the final product. Design is multidisciplinary and has a significant influence on other processes, as well as on the final product, in terms of quality and value (Koskela, 2000; Macmillan et al., 2001; Tzortzopoulos et al., 2001).

There are some models developed to represent building design process. In the traditional RIBAs framework, the initial stage of the process of designing a building project is called Preparation, which is composed by appraisal ("identification of client's needs and objectives") and design brief ("development of initial statement of requirements"). The second stage is Design, which is composed by conceptual design, design development, and technical design, followed by Pre-construction, Construction and Use stages (RIBA, 2007). In more recent times, demolition and design for deconstruction were included in the design agenda (Morgan and Stevenson, 2005).

An expressive part of value, cost, and waste generated in the life cycle of a building is defined in the initial phases of design. In some cases early decision in design define from 70% to 80% of the final cost (Bouchlaghem et al., 2006; Rafiq et al., 2005) and have considerable impact on the building performance (Wang et al., 2005).

Design process generated a large quantity of knowledge (Baldwin et al., 1999; Langford and Retik, 1996; Yusuf and Alshawi, 1999). The success of the final design solution depends on how design team can coordinate knowledge at the earliest possible time, after Meniru et al. (2003). However, time constraints and inadequate communication between client and architect and among the design team have negative influence on design quality (Ballard and Koskela, 1998; Kamara et al., 2000; Luck and McDonnell, 2006; Yu et al., 2005). Batty (1995) says that the lack of time available can have a number of consequences, including a lack of desire, on the part of the design team, in terms of taking risks with new materials or systems.

To identify properly user requirements in the early stage of design reduces the effort of re-design in subsequent stages. A good time to make decision is during briefing. Briefing is an initial step of designing process and often the most important, since it has to establish project goals and set a basis to develop conceptual design (Peña and Parshall, 2001). It's also known as architectural programming. Briefing was not needed in other times, when buildings were more simples. Since industrial revolution, buildings are increasingly specialized, requiring elaborate and specific briefs (Donia, 1998). Barrett et al. (1996) define briefing as a systemic process by which client/user's ideas are made explicit and formalized. In general, briefing is also placed as a planning activity in the building design process.

Early approaches considered brief as a static document, produced at a specific point of time. However, authors argue that briefing must be dynamic and practiced throughout the design process (Aouad et al., 1998; Barrett et al., 1996; Tzortzopoulos et al., 2006). For instance, in the RIBA's Plan of Work, brief evolves since an "initial brief" until a "detailed project brief" and this evolution occurs in parallel with the development of conceptual and scheme design (Kamara et al., 2001; RIBA, 2007). Kamara et al. (2001) suggest that the briefing may involve two parts: a strategic program and the brief itself.

2.2. Knowledge-based approaches in design

Since briefing evolves and design process deal with a large amount of knowledge, it became important to provide designers with tools to manage knowledge. Some ways explored to improve briefing and manage design knowledge are value management (Kelly et al., 2005; Yu et al., 2005), QFD (Kamara et al., 1999), planning design with Last Planner (Koskela et al., 1997; Tzortzopoulos et al., 2001), and concurrent engineering or collaborative schemes of design (Bouchlaghem et al., 2006; Kamara and Anumba, 2000; Kamara et al., 2001; Koskela and Huovila, 1997; Macmillan et al., 2001; Marir et al., 2000; Meniru et al., 2003; Yan-chuen et al., 2000).

Other way to improve design is through Decision Support Systems (DSS). DSS are interactive, computer-based systems that help decision-makers to solve structured or unstructured problems, which have multiple attributes, objectives or goals (Power, 2002). It has been used basically two types of DSS in design: Expert Systems (also known as Knowledge-Based systems or Knowledge-Driven Decision Support Systems), and systems using Case-Based Reasoning or other Data Mining tools.

An Expert System has a knowledge base (KB) and an inference engine. A KB is a collection of organized knowledge, rules, and procedures. The active component is the inference engine, which contain rules elicited from a domain expert. It is built using explicit, structured knowledge (Power, 2002). Some ES has been proposed in design, with applications on building refurbishment (Kaklauskas et al., 2005; Zavadskas et al., 2006), semi-automated design of housing (González-Uriel and Roanes-Lozano, 2004), and housing evaluation (Natividad-Jesus et al., 2007).

More focus has been also given to DSS using Case-Based Reasoning (CBR). CBR is an usual tool to reasoning and learning with successful applications in several fields (Watson, 1997). This technique considers solutions for new problems using as base solutions adopted for previous problems (described as “cases”), identified in a data base through a mechanism of case selection based on the similarity among problem case and each case available in database (Aamodt and Plaza, 1994; Kolodner, 1993; Watson, 1997). CBR have some advantages upon ES, such as it does not need explicit models to obtain problem solution and it has flexibility to work with great amounts of data. A further advantage is that CBR has the possibility to learn with new cases, being easy to keep the application updated. A weakness of CBR is to adjust selected cases to perform numerical results (Watson, 1997). CBR has been applied in design since the end of 80s (Maher, 1987; Pearce et al., 1992).

There are some applications of CBR in briefing. SEED-Pro search for client needs, budget and constraints, generating an architectural program (Akin et al., 1995; Donia, 1998). Marir et al. (2000) describes a CBR system designed to improve the specification of construction projects that involves the integration of information and supports concurrent engineering and decision making for the effective management and realization of all stages of a construction project's lifecycle. Van Leeuwen et al. (2000) propose a system for housing refurbishment, which uses knowledge about architectural design, cost, and building products. The system considers user requirements through a CBR system that have a case base with a number of typical housing layouts. Serpell and Rueda (2007) developed a CBR system to briefing, using a three step process that search for similar cases into definition of a new project.

2.3. Knowledge discovery in real estate market

Although the literature presents several efforts to improve the building design process, most of the examples consider designer knowledge (managing design/team knowledge or using previous experiences, expressed in cases or rules elicited from experts).

There is relatively little effort in applications to discover building requirements based on user knowledge.

However, it is possible to identify the preferences of building customers from real estate market information, regarding the price paid for each property. Since customers' decisions are mainly economically rational, prices are proportional to the quality level perceived or, in a larger sense, the quality level perceived by users. In other words, price is a proxy variable to product' value.

Properties may be considered as multi-dimensional commodities, considering the simultaneous influence of several characteristics that form the final price. Thus, properties are heterogeneous goods and have a unique bundle of attributes. They differ in terms of design, size, inner configuration, construction quality, and location, for example. Therefore, there is a great variety of products in real estate market, and their heterogeneity makes the direct comparison a difficult task. As a consequence, initially it is hard to understand the relative importance of each property characteristic, with respect of its participation in the final price. It is only known the total price (Harvey, 1996; Lavender, 1990; Robinson, 1979).

Hedonic price models (HPM) search to establish the relationship between the property price and its characteristics. In hedonic models, the goods are described through a "bundle of attributes", congregating the characteristics that are important. As the partial prices related to each attribute cannot be directly isolated, because it does not have specific markets for each one, the prices are obtained indirectly. The implicit prices of each one of these attributes, also called hedonic prices or "shadow prices", are the prices related with each attribute of the property. By regressing the characteristics of the building on the observed price, it can be extracted the contribution of each attribute on the total price. The tool generally used to obtain the coefficients that measure these contributions is multiple regression analysis (MRA), a well known statistical tool, used in almost all hedonic studies. Following the hedonic price theory, the coefficients represent the prices that the purchasers are willing to pay, on average, for these characteristics (Rosen, 1974; Sheppard, 1999).

To give effective support to designers, these models must be sustainable, meaning that they might be based on sound data. The analysis of large databases is seen as a typical challenge to be approached by knowledge discovery in databases (KDD). KDD is a relatively novel approach to data analysis. It consists of a special data organization and techniques to allow the revelation of knowledge that presumably is occult in the data. This area appeared in the end of 80's, as an alternative for the analysis of very large databases (Fayyad et al., 1996).

The knowledge discovery occurs through different phases, composed by three basic stages of the process: preprocessing, selection of relevant data, and data mining. The preprocessing phase includes data collection and preparation. It may use several statistical techniques, such as clustering, multiple regression analysis, factor analysis, sampling, and descriptive statistics. The output of this stage is a reliable database. The second stage is data selection, looking for relevant data for each problem. It uses sampling, CBR or clustering, for instance. In the data mining phase other techniques may be chosen to solve the knowledge discovery problem, such as neural networks, multiple regression analysis, clustering, case-based reasoning, genetic algorithms and fuzzy rule-based systems (Berry and Linoff, 2000; Hair et al., 1998; Pyle, 1999). There are some studies in construction field, investigating delay in construction projects (Kim et al., 2008; Soibelman and Kim, 2002).

3. Methodology and data

3.1. Proposed system

This study considers Knowledge Discovery in Databases (KDD) paradigm in a hybrid system, using Case-Based Reasoning (CBR) and Multiple Regression Analysis (MRA) for the construction of hedonic price models. Investigation presented was based in a simulation using actual real estate data.

KDD was included because it permits manage large databases. CBR is a convenient way to select properties, using a similarity index to find relevant properties (those are similar to the building in study). As CBR has difficulties to manage and to adapt numerical data, MRA is used to generate quality indexes, based on the relationship of property price and its characteristics. This kind of relationship is presented in equation form (a hedonic price model).

However, to develop studies using real estate market it is need to include two additional steps in the basic scheme of KDD/CBR. Detailed samples are necessary to generate specific price models. As properties in general have a large set of important attributes, detailing all cases in a large database since first data collecting is not viable. The solution for this apparent paradox is to collect basic information for all cases (sufficient to permit case selection in CBR) and to complement data sample after the selection, as well as, to make total collecting work only for selected properties (sufficient to generate useful HPM).

In traditional CBR' scheme some similar cases are selected which will be adapted to provide a solution. In this view, the final result is provided by CBR itself. The second difference in our proposal is that the system to select a sample to be modeled, using an additional tool to generate the result (MRA). Then, the system proposed is composed by five stages:

- (1) To create database - to collect and preprocess data from real estate market;
- (2) To select a sample with relevant data through CBR;
- (3) To enrich data (detailing the sample);
- (4) To generate HPM using MRA;
- (5) To simulate value of different alternatives in design process.

Figure 1 shows the general configuration of proposed DSS.

Figure 1. Configuration of the proposed system

Figura 1. Configuración del sistema propuesto

As follow, this system is demonstrated trough an empirical study consisting of the estimation and use of a price model considering data from middle-class segment apartments in Porto Alegre, a Southern Brazilian city. The first four stages are presented in this section and the fifth is on the next section.

3.2. Creating a database

This stage of the study is based on similar research about the real estate modeling (Author, 2006). Initially, it had been collected data about local real estate market, forming a database composed by more than 30,000 cases. The information about these units was obtained from Sales Tax files in the Porto Alegre Tax Department.

In this phase, it was collected basic data, such as sales price (as declared by taxpayer), privative and total floor area, year of the building completion, and the construction quality level. In sequence, it was collected variables indicative of the location quality, such as the distance from the central business district and from the main centers of commerce and leisure. It was also collected qualitative variable indicating the quality of the neighborhood of the buildings. After the data collection phase, preprocessing operations were conducted, resulting in a data set ready for general modeling.

3.3. Selecting data

It was simulated a design process of apartments on pre-sales market. The relevant data for the segment studied was selected through CBR, considering apartment units in the medium class regions of the city. Selection was made using basic case attributes, such as sales prices, floor area and location. It was used a similarity mechanism based on nearest neighborhood algorithm (K-NN). At all, it was formed a dataset composed by 110 apartment units.

3.4. Enriching data

The sample was detailed by the use of sources such as design plans and building photos. Important information of this market segment was identified, as number of bedrooms and bathrooms; existence of fireplace, barbecue equipment, balcony and laundry; number of privative parking spaces; characteristics of the building leisure (presence of swimming pool, sauna, spaces for sports and playground), as well as the quality of the construction.

The region considered is relatively uniform, with respect to accessibility and neighborhood quality. The sample is composed of new properties, and prices were relatively stable in the period of data collection. Therefore, it was not necessary to include variables such as age, location, neighborhood, and time of sale.

3.5. Generating HPM

To generate the hedonic price modeling, different quantitative variables were considered: the sale price (€), the floor area (m²), the number of simple bedrooms, the number of bedrooms with an exclusive bathroom, the number of leisure spaces (including swimming pool/sauna/fitness center/playground), the number of privative parking spaces and the assessment of building quality (from 1 – very poor to 10 – very good).

Dummy variables were regarded, with respect to the presence of the follow elements: bathtub, home office, balcony with barbecue equipment, simple balcony, fireplace, and laundry space. The dummy variable were considered equals to one if the property has the presence of the attribute, and zero otherwise.

Equation 1 presents the basic model, while Table 1 shows the descriptive statistics:

$$SP = a_0 + a_1*FA + a_2*BBr + a_3*SBr + a_4*BT + a_5*HOr + a_6*BBe + a_7*SBa + a_8*FP + a_9*LA + a_{10}*LE + a_{11}*PPS + a_{12}*BQ + \varepsilon \quad (1)$$

where:

- *SP* is the sale price in Euros;
- *FA* is the floor area in squared meters (m²);
- *BBr* is the number of bedrooms with an exclusive bathroom;
- *SBr* is the number of simple bedrooms;
- *BT* is equals to one if the property has a bathtub and zero otherwise;
- *HOr* is equals to one if the property has a home office in a separated room and zero otherwise;

- *BBE* is equals to one if the property has a balcony with barbecue equipment and zero otherwise;
- *SBa* is equals to one if the property has a simple balcony and zero otherwise;
- *FP* is equals to one if the property has a fireplace and zero otherwise;
- *LA* is equals to one if the property has a laundry space and zero otherwise;
- *LE* is the number of leisure spaces (including swimming pool/sauna/fitness center/playground);
- *PPS* is the number of privative parking spaces;
- *BQ* is the assessment of building quality;
- a_i are the coefficients to be estimated; and
- ϵ is the stochastic term.

Table 1. Summary statistics of the quantitative variables**Tabla 1.** Resumen estadístico de las variables cuantitativas

<i>Variables</i>	<i>Unity/ Unidad</i>	<i>Maximum value/ Valor máximo</i>	<i>Minimum value/ Valor mínimo</i>	<i>Mean/ Promedio</i>	<i>Standard deviation/ Desviación estándar</i>
Sale Price (SP)/ Precio de venta (SP)	Euros	364,000.00	49,142.85	126,889.08	49,469.58
Floor Area (FA)/ Superficie (FA)	m ²	287.77	39.32	104.22	41.81
Bedroom with Bathroom (BBr)/ Dormitorio con baño (BBr)	-	3	0	0.60	0.61
Simple Bedroom (SBr)/ Dormitorio simple (SBr)	-	3	0	1.87	0.65
Simple Balcony (SBa)/ Terraza simple (SBa)	-	2	0	0.64	0.57
Leisure (LE)/ Recreación (LE)	-	4	0	0.82	1.02
Private Parking Space (PPS)/ Estacionamiento privado (PPS)	-	3	0	0.91	0.83
Building Quality (BQ)/ Calidad de la construcción (BQ)	-	10.0	4.5	5.83	0.78

While the majority of these attributes is commonly used in hedonic studies (Ball, 1973; Boyle and Kiel, 2001; Chau and Wong, 2004; Chau et al., 2001; Din et al., 2001; Smith et al., 1988), it may be considered the local cultural differences. For instance, the importance of balconies was demonstrated by Chau and Wong (2004).

In the present case, the attribute “balcony with barbecue equipment” (BBE) means an important characteristic for the real estate market in Porto Alegre. The local culture privileges the consumption of meat baked in meetings with family and friends. The inclusion of barbecue equipment in upper-middle market segment apartments in the city of Porto Alegre was initiated about 30 years ago, and it was quickly spread for other kinds of properties and also on surrounding ones. It consists of a strong sales appeal, and for many cases, the design of the privative area is defined using this element as principal start point.

Also, the presence of a home office in a separated room (HOr) is considered a very important space in this segment. By another hand, the building leisure was represented in one variable, counting the presence of different elements (swimming pool, sauna, fitness center, and playground), based on results from initial studies that indicated that the elements individually had no statistical importance.

4. Results and discussion

4.1 Statistical analysis of HPM

The regression results from the sample of the data selected are shown in Table 2. It should be mentioned that some other models have been developed, and linear specification provided better results.

Table 2. Regression results - linear model
Tabla 2. Resultados del análisis de regresión – modelo lineal

Variable	Coefficient/ Coeficiente	t-statistic/ t-estadístico
Constant/ Constante	-173,428.068	-19.141
Floor Area (FA)/ Superficie (FA)	340.530	10.418
Bedrooms with Bathroom (BBr)/ Dormitorios con baño (BBr)	16,706.302	7.113
Simple Bedroom (SBr)/ Dormitorio simple (SBr)	4,565.228	2.378
Bathtub (BT)/ Tina de baño (BT)	7,063.530	2.137
Home Office (HOr)/ Escritorio (HOr)	14,659.450	4.070
Balcony with Barbecue Equipment (BBE)/ Terraza con barbacoa (BBE)	23,943.206	10.590
Simple Balcony (SBa)/ Terraza simple (SBa)	6,536.118	3.393
Fireplace (FP)/ Chimenea (FP)	17,247.389	6.969
Laundry Space (LA)/ Espacio para lavadero (LA)	5,876.922	2.021
Leisure Space (LE)/ Espacio para recreación (LE)	4,388.420	4.438
Private Parking Space (PPS)/ Estacionamiento privado (PPS)	8,778.206	6.091
Building Quality (BQ)/ Calidad de construcción (BQ)	34,994.143	26.390
Adjusted R²/ R² ajustado		0.9694
Number of observations/ Número de observaciones		110
Dependent variable/ Variable dependiente		Sale Price (SP)/ Precio venta (SP)

The equation of the HPM, regarding the variables and the coefficients from Table 2 is presented in Equation 2:

$$SP = -173,428.068 + 340.530(FA) + 16,706.302 (BBr) + 4,565.228 (SBr) + 7,063.530 (BT) + 14,659.450 (HOr) + 23,943.206 (BBE) + 6,536.118 (SBa) + 17,247.389 (FP) + 5,876.922 (LA) + 4,388.420 (LE) + 8,778.206 (PPS) + 34,994.143 (BQ) \quad (2)$$

As presented in Table 2, all variables have shown importance and coefficients are significant at 5 percent level or better. The results indicate a high adjusted R² of 0.9694. The analysis of graphs with collected prices x estimated prices, and estimated prices x standardized errors, although are not shown here, did not indicate the presence of trends or other statistical problems, such as autocorrelation, outliers, not constant error variance, etc. Likewise, the correlation analysis did not point out strong correlations among the explanatory variables, since they did not exceed 0.6. Therefore, it may be concluded that the model presents adequate statistical conditions and can be helpful for market analysis, in terms of value estimation.

The analysis of the coefficients set allows the comparison among variables and the identification of alternatives with a major aggregated value. For example, results indicate that a simple bedroom (SBr) or a collective space for gymnastics, included in the leisure spaces (LE), adds similar values for the users in the total price of the property. In this example, the hedonic price of SBr is 4.5 thousand Euros and the hedonic price of LE is 4.4 thousand Euros. On the other hand, the difference between the price variation considering simple bedroom (SBr, 4.5 thousand Euros) and a bedroom equipped with a bathroom (BBR, 16.7 thousand Euros) certainly would justify the inclusion of the bathroom as a better alternative.

However, in other cases, the importance of some attributes is not too clear. For example, the hedonic coefficients of some variables such as fireplace (FP), bathtub (BT), and balcony with barbecue equipment (BBE) represent not exclusively the named equipment, but actually a larger set of elements. In fact, they are emblematic elements and work as value signs, indicating a superior category of the property. They have also strong influence on design. For instance, the inclusion of a balcony or a fireplace causes the increasing and redistribution of inner spaces, and a bathtub causes the magnifying of the bathroom area. Most of the time, it is very difficult to add those elements, during the production phase. After the conclusion of the production, it is more difficult, or even impossible. Designers should consider those elements since the first drafts, reinforcing the importance of the knowledge to support the early design stages.

In most of the properties of the sample, balconies with barbecue equipment (BBE) are relatively large spaces, from about 8 to 15m² and they also play as an extra living/dining room. Simple balconies (SBa) are located, mostly, in the bedrooms and have distinct functions: to promote the contact with the exterior space and to improve the natural ventilation. In general, they are smaller, around 6m². From a comparison of both hedonic prices (BBE = 23.9 thousand Euros; SBa = 6.5 thousand Euros) it is possible to conclude that simple balconies are less intense used and have minor importance for the users.

4.2. Simulation using information from HPM

When a designer is faced to a new project composed by several 100 m² apartment units, for example, different configurations are possible. In this case, the value of distinct elements should be helpful to guide the designer, in order to maximize the user preferences and the final price of the project. In this context, as follow it is presented a demonstration of two situations, using the results obtained in the HPM generated. Both examples consist of a 100m² apartment with two parking spaces, regarding a building leisure composed of swimming pool and playground, with a building quality level equals to seven.

Situation A: a living room (24m²), one bedroom with bathroom (20m²), two bedrooms (19,8m²) two bathrooms (5,2m²), a balcony (4,5m²), kitchen and services (10m²), corridor (4m²), considering the area of walls equals to 12,5m².

Situation B: a living room connected to a balcony with barbecue equipment (31,7m²), two bedrooms with bathrooms (35,7m²), a bathroom (1,3m²), home office (4m²), kitchen (8m²), laundry space (3m²), corridor (4m²), and area of walls equals to 12,3m².

Table 3 shows the description of the two situations, based on the variables considered in the HPM.

Table 3. Variable levels on situations A and B
Tabla 3. Nivel de las variables para los casos A y B

Variable	Situation A/ Caso A	Situation B/ Caso B
Floor area (m ²) (FA)/ Superficie (m ²) (FA)	100	100
Bedrooms with bathroom (units) (BBr)/ Dormitorios con baño (unidades) (BBr)	1	2
Simple bedrooms (units) (SBr)/ Dormitorios simples (unidades) (SBr)	2	0
Home office (yes = 1; no = 0) (HOr)/ Escritorio (si = 1; no = 0) (HOr)	0	1
Balcony with barbecue equipment (yes = 1; no = 0) (BBE)/ Terraza equipada con barbacoa (si = 1; no = 0) (BBE)	0	1
Simple balcony (yes = 1; no = 0) (SBa)/ Terraza simple (si = 1; no = 0) (SBa)	1	0
Laundry space (yes = 1; no = 0) (LA)/ Espacio para lavadero (si = 1; no = 0) (LA)	0	1
Leisure spaces (yes = 1; no = 0) (LE)/ Espacios para recreación (si = 1; no = 0) (LE)	2	2
Privative parking spaces (units) (PPS)/ Espacios para estacionamiento privado (unidades) (PPS)	2	2
Building quality (from 1 to 10) (BQ)/ Calidad de la construcción (de 1 a 10) (BQ)	7	7

Equations 3 and 4 shows the equations of Situation A and B, as well as and the sale price resulted from both.

Equation of Situation A

$$SP = -173,428.068 + 340.530 (100) + 16,706.302 (1) + 4,565.228 (2) + 6,536.118 (1) + 4,388.420 (2) + 8,778.206 (2) + 34,994.143 (7) \quad (3)$$

$$SP = \text{€ } 164,290.02$$

Equation of Situation B

$$SP = -173,428.068 + 340.530 (100) + 16,706.302 (2) + 4,659.450 (1) + 23,943.206 (1) + 5,876.922 (1) + 4,388.420 (2) + 8,778.206 (2) + 34,994.143 (7) \quad (4)$$

$$SP = \text{€ } 199,808.94$$

According to Equations 3 and 4, Situation B shows higher sale price estimation, compared to the sale price of Situation A. Since both situations may have approximately the same building costs because of the same area, building quality, and same leisure, Situation B shows a significant financial and economical advantage. The difference of both sale prices is around 35 thousand Euros.

Also, regarding the fact that Situation B strongly regards the market preferences, because of the real state data used, it is expected less refurbishment, rebuilding, and demolition during the life cycle of the building. For example, to transform the apartment of Situation A to the apartment of Situation B, it would be necessary to remove something like 13m², generating around 2m³ of construction and demolition waste amount.

One may conclude that the hedonic model (Equation 2) have good potential to be used in design. They can be used to support decisions in the early design phase, being helpful in trade-offs situations, making it possible to evaluate the impact of different configuration decisions on the final sale price. By this way, it can be used for estimate values and designers can simulate the value of some alternatives, deciding for that ones that result in greater market value.

5. Conclusions

This study proposes a novel approach, since it suggests the use of market real estate data to discover user' requirements. It is a hybrid KDD system, using the techniques of CBR for case selection and multiple regression analysis for the generation of price models. The principle behind the system is that there is occult knowledge about user requirements in sales data, which maybe discovered through hedonic price models.

One application was presented, based on actual property data from a Brazilian city. The hedonic model calculated had a good statistical performance.

As main advantages, five points can be highlighted. Firstly, it is a numerical model that allows simulation and designer can study a mix of different options, exploring optimal or best solution in terms of the main user' requirements. Secondly, the model can be obtained through actual real estate market data. Thirdly, the information obtained might be used to support decision in the very early stage of design process, when most of the total cost and price of the project are defined. Fourthly, the use of automated tools may improve the efficiency of the process, since the results are easily and quickly obtained if the database is organized. Finally, HPM identify economically viable choices, since the model are based on actual market data.

In summary, hedonic price models can be understood as a sound way to identify requirements and preferences of buildings users through an indirect form. Thus, the phase of conception of new products can become more objective and intelligent.

6. Referencias/ References

- Aamodt A. and Plaza E. (1994)**, "Case-based reasoning: Foundational issues, methodological variations, and system approaches", *AI Communications*, Vol. 7 No. 1, pp. 39-59. - <http://iospress.metapress.com/index/316258107242JP65.pdf>
- Akin Ö., Sen R., Donia M. and Zhang Y. (1995)**, "SEED-Pro: Computer Assisted Architectural Programming in SEED", *Journal of Architectural Engineering*, Vol. 1 No. 4, pp. 153-161. - available at <http://web.ebscohost.com/ehost/detail?sid=35686709-ec78-4d5d-b7f9-bb86ef0e518e%40sessionmgr113&vid=1&hid=122&bdata=Jmxhbm9cHQYtYnI2OT1laG9zdC1saXZl#db=a9h&AN=6783082>
- Aouad G., Hinks J., Cooper R., Sheath D.M., Kagioglou M. and Sexton M. (1998)**, "An IT Map for a generic design and construction process protocol" *Journal of Construction Procurement*, Vol. 4 No. 1, pp. 132-151. - available at <http://eprints.lancs.ac.uk/39899/>
- Baldwin, A.N., Austin, S.A., Hassan, T.M. and Thorpe, A. (1999)**, "Modeling information flow during the conceptual and schematic stages of building design", *Construction Management and Economics*, No. 17, pp.155-167. - available at <http://www.tandfonline.com/doi/abs/10.1080/014461999371655>
- Ball M.J. (1973)**, "Recent empirical work on the determinants of relative house prices", *Urban Studies*, Vol. 10 No. 2, pp. 213-233. -

available at <http://usj.sagepub.com/content/10/2/213.short>

- Ballard G. and Koskela L. (1998)**, "On the agenda for design management research", in Proceedings of the 6th Annual Conference of the International Group for Lean Construction, Guarujá Beach. - available at <http://www.ce.berkeley.edu/~tommelein/IGLC-6/BallardAndKoskela.pdf>
- Barrett P.S., Hudson J. and Stanley C. (1996)**, "Is briefing innovation?", in Langford, D.A. and Retik, A. (Eds.), The Organization and Management of Construction, Shaping Theory and Practice - Volume 3, Managing Construction Information, E & FN Spon, London, pp. 87-95.
- Batty W.J. (1995)**, "The development of interactive strategic design environments", Pattern Recognition, Vol. 28 No.10, pp.1477-1481. - available at <http://www.sciencedirect.com/science/article/pii/003132039400152C>
- Berry M.J. and Linoff G. (2000)**, Mastering Data Mining, Wiley Computer Publishing, New York, NY.
- Bouchlaghem D., Shang H., Anumba C.J., Cen M., Miles J. and Taylor M. (2006)**, "ICT-enabled collaborative working environment for concurrent conceptual design", Architectural Engineering and Design Management, Vol. 1, pp.261-280. - available at <http://www.tandfonline.com/doi/abs/10.1080/17452007.2005.9684597>
- Boyle M.A. and Kiel K.A. (2001)**, "A survey of house price hedonic studies of the impact of environmental externalities", Journal of Real Estate Literature, Vol. 9 No. 2, pp.117-144. - available at <http://ares.metapress.com/index/23U082061Q53QPM3.pdf>
- Chau K.W. and Wong S.K. (2004)**, "The value of the provision of a balcony in apartments in Hong Kong", Property Management, Vol. 22 No. 3, pp. 250-263. - available at <http://www.emeraldinsight.com/journals.htm?articleid=845575&show=abstract>
- Chau K.W., Ma V.S. M. and Ho D.C. W. (2001)**, "The pricing of 'Luckiness' in the apartment market", Journal of Real Estate Literature, Vol. 9 No. 1, pp. 31-40.
- Din A., Hoesli M. and Bender A. (2001)**, "Environmental variables and real estate prices", Urban Studies, Vol. 38 No. 11, pp. 1989-2000. - available at <http://usj.sagepub.com/content/38/11/1989.short>
- Donia M. (1998)**, Computational Modeling of Design Requirements for Buildings, Ph.D. Dissertation, School of Architecture, Carnegie Mellon University, Pittsburgh, PA. - available at <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.125.1743&rep=rep1&type=pdf>
- Fayyad U.M., Piatetsky-Shapiro G., and Smith P. (1996)**, "From data mining to knowledge discovery: An overview", in Fayyad, U.M., Piatetsky-Shapiro, G., Smyth, P. and Uthurusamy, R. (Eds.), Advances in Knowledge Discovery and Data Mining, AAAI Press/MIT, Menlo Park, CA/Cambridge, MA, pp. 1-34.
- González M.A.S. and Formoso C.T. (2006)**, "Mass appraisal with genetic fuzzy rule-based systems", Property Management, Vol. 24 No. 1, pp. 20-30. - available at <http://www.emeraldinsight.com/journals.htm?articleid=1541555&show=abstract>
- González-Uriel A. and Roanes-Lozano E. (2004)**, "A knowledge-based system for house layout selection", Mathematics and Computers in Simulation, Vol. 66 No. 1, pp. 43-54. - available at <http://www.sciencedirect.com/science/article/pii/S0378475404000771>
- Hair Jr. J. F., Anderson R.E., Tatham R.L. and Black W.C. (1998)**, Multivariate Data Analysis, 5ed., Prentice-Hall, Upper Saddle River, NJ.
- Harvey, J. (1996)**, Urban Land Economics, 4ed., MacMillan, London.
- Juan Y.K., Shih S.G. and Perng Y.H. (2006)**, "Decision support for housing customization: A hybrid approach using case-based reasoning and genetic algorithm", Expert Systems with Applications, Vol. 31 No. 1, pp. 83-93. - available at <http://www.sciencedirect.com/science/article/pii/S0957417405001910>
- Kaklauskas A., Zavadskas E.K. and Raslanas S. (2005)**, "Multivariate design and multiple criteria analysis of building refurbishments", Energy and Buildings, Vol. 37 No. 4, pp. 361-372. - available at <http://www.sciencedirect.com/science/article/pii/S0378778804002385>
- Kamara J.M., Anumba C.J., Members A. and Evbuomwan N.F.O. (1999)**, Client requirements processing in construction: a new approach using QFD. Journal of architectural engineering, Vol. 5 No.1, pp. 8-15.
- Kamara J.M. and Anumba C.J. (2000)**, "Client requirements processing for concurrent life-cycle design and construction", Concurrent Engineering, Vol. 8 No.2, pp.74-88. - available at <http://cer.sagepub.com/content/8/2/74.short>
- Kamara J.M., Anumba C.J. and Evbuomwan N.F.O. (2000)**, "Process model for client requirements processing in construction", Business Process Management Journal, Vol. 6 No. 3, pp. 251-379. - available at <http://www.emeraldinsight.com/journals.htm?articleid=843453&show=abstract>
- Kamara J.M., Anumba C.J. and Evbuomwan N.F.O. (2001)**, "Assessing the suitability of current briefing practices in construction within a concurrent engineering framework", International Journal of Project Management, Vol. 19, pp.337-351. - available at <http://www.sciencedirect.com/science/article/pii/S0263786300000156>
- Kelly J., Hunter K., Shen G. and Yu A. (2005)**, "Briefing from a facilities management perspective", Facilities, Vol. 23 No. 7/8, pp.356-367. - available at <http://www.emeraldinsight.com/journals.htm?articleid=1502622&show=abstract>
- Kim H., Soibelman L. and Grobler F. (2008)**, "Factor selection for delay analysis using knowledge discovery in databases", Automation in Construction, Vol. 17 No. 5, pp. 550-560. - available at <http://www.sciencedirect.com/science/article/pii/S0926580507001306>
- Kolodner J.K. (1993)**, Case-Based Reasoning, Morgan Kaufmann Publishers, San Francisco.

- Koskela L. (2000)**, An Exploration Towards a Production Theory and its Application to Construction, Ph.D. Dissertation, Technical Research Centre of Finland – VTT, Helsinki. - available at <http://mail.leanconstruction.org/pdf/P408.pdf>
- Koskela L. and Huovila P. (1997)**, “On Foundations of concurrent engineering”, in Anumba, C. and Egbuomwan, N. (Eds.). Concurrent Engineering in Construction (CEC97), The Institution of Structural Engineers, London, pp. 22-32.
- Koskela L., Ballard G. and Tanhuanpaa V.-P. (1997)**, “Towards lean design management”, in Proceedings of the 5th Annual Conference of the International Group for Lean Construction, Gold Coast, Australia, pp.1-13. - available at <http://www.mendeley.com/research/towards-lean-design-management/>
- Langford D.A. and Retik A. (Eds.) (1996)**, The Organization and Management of Construction, Shaping Theory and Practice, Vol. 3 - Managing Construction Information, E & FN Spon, London.
- Lavender S.D. (1990)**, Economics for Builders and Surveyors, Longman, Essex, UK.
- Luck R. and McDonnell J. (2006)**, “Architect and user interaction: The spoken representation of form and functional meaning in early design conversations”, Design Studies, Vol. 27 No. 2, pp.141-166. - available at <http://www.sciencedirect.com/science/article/pii/S0142694X05000694>.
- Macmillan S., Steele J., Austin S., Kirby P. and Spence R. (2001)**, “Development and verification of a generic framework for conceptual design”, Design Studies, Vol. 22 No. 2, pp. 169-191. - available at <http://www.sciencedirect.com/science/article/pii/S0142694X00000259>
- Maher M. L. (1987)**, “Expert Systems for structural design”, Journal of Computing in Civil Engineering, Vol. 1 No. 4, pp. 270-283.
- Marir F., Rezgui Y. and Benhadj-Djilali R. (2000)**, “A case-based expert system for construction project process activity specifications”, International Journal of Construction Information Technology, Vol. 8 No. 1, pp.53-73. - available at <http://eprints.kingston.ac.uk/3202/>
- Meniru K., Rivard H. and Bédard C. (2003)**, “Specifications for computer-aided conceptual building design”, Design Studies, Vol. 24, pp.51-71. - available at <http://www.sciencedirect.com/science/article/pii/S0142694X02000091>
- Morgan C. and Stevenson F. (2005)**, Design and Detailing for Deconstruction - SEDA Design Guide for Scotland - No.1, Scottish Ecological Design Association (SEDA), Northumberland, UK. - available at http://www.bot.yildiz.edu.tr/ids09/_data/_readings/DESIGN%20AND%20DETAILING%20FOR%20DECONST.pdf
- Natividade-Jesus E., Coutinho-Rodrigues J. and Antunes C.H. (2007)**, “A multicriteria decision support system for housing evaluation”, Decision Support Systems, Vol. 43 No. 3, pp. 779-790. <http://www.sciencedirect.com/science/article/pii/S0167923606002156>
- Pearce M, Goel A.K., Kolodner J.L., Zimring C., Sentosa L. and Billington R. (1992)**, “Case-based design support: A case study in architectural design”, IEEE Expert, Vol. 7, pp. 14-20. - available at http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=163668
- Peña W. M. and Parshall S. A. (2001)**, Problem Seeking: An Architectural Programming Primer, 4ed., John Wiley and Sons, New York, NY. - available at <http://www.sciencedirect.com/science/article/pii/S0142694X05000517>
- Power D. J. (2002)**, Decision Support Systems: Concepts and Resources, Quorum, Westport.
- Pyle D. (1999)**, Data Preparation for Data Mining, Morgan Kaufmann, San Francisco, CA.
- Rafiq M.Y., Mathews J. D. and Bullock G. N. (2003)**, “Conceptual Building Design - Evolutionary Approach”, Journal of Computing in Civil Engineering, Vol. 17 No. 3, pp. 150-158. - available at http://ascelibrary.org/cpo/resource/1/jccee5/v17/i3/p150_s1
- Robinson R. (1979)**, Housing Economics and Public Policy, MacMillan, London.
- Rosen S. (1974)**, “Hedonic prices and implicit markets: product differentiation in pure competition”, Journal of Political Economy, Vol. 82 No. 1, pp. 34-55. - available at <http://www.jstor.org/sici?sici=0022-3808%28197401%2F02%2982%3A1%3C34%3AHPAIMP%3E2.0.CO%3B2-U&>
- Royal Institution of British Architects (RIBA) (2007)**, “RIBA outline plan of work”, available at: <http://www.architecture.com/Files/RIBAProfessionalServices/ClientServices/>.
- Serpell A.F. and Rueda J.M. (2007)**, “Modeling a project scope using a Case-Based Reasoning approach”, in Proceedings of the CIB W78 24th International Conference, CIB, Maribor, Slovenia, pp.329-335.
- Sheppard S. (1999)**, “Hedonic analysis of housing markets”, in Cheshire, P. C., and Mills, E. S. (Eds.), Handbook of Applied Urban Economics - Vol. 3, Elsevier, New York, NY, pp.1595-1635.
- Shin Y., An S.H., Cho H.H., Kim G.H. and Kang K.I. (2008)**, “Application of information technology for mass customization in the housing construction industry in Korea”, Automation in Construction, Vol. 17 No. 7, pp. 831-838.
- Smith L.B., Rosen K.T. and Fallis G. (1988)**, “Recent developments in economic models of housing markets”, Journal of Economic Literature, Vol. 26 No. 1, pp. 29-64.
- Soibelman L. and Kim H. (2002)**, “Data preparation process for construction knowledge generation through knowledge discovery in databases”, Journal of Computing in Civil Engineering, Vol. 16 No. 1, pp.39-48. - available at <http://link.aip.org/link/?JCCEE5/16/39/1>
- Tzortzopoulos P., Cooper R., Chan P. and Kagioglou M. (2006)**, “Clients’ activities at the design front-end”, Design Studies, Vol. 27, pp.657-683. - available at <http://www.sciencedirect.com/science/article/pii/S0142694X06000445>

- Tzortzopoulos P., Formoso C.T. and Betts M. (2001)**, "Planning the product development process in construction: an exploratory case study", in Proceedings of the 9th Annual Conference of the International Group for Lean Construction, National University of the Singapore, Singapore, pp. 103-116. - available at <http://cic.vtt.fi/lean/singapore/Tzortzopoulosetal.pdf>
- Van Leeuwen J.P., de Vries B. and van den Oetelaar E. (2000)**, "A decision support system for building refurbishment design", in Proceedings of the CIB W78 International Conference on Construction Information Technology (CIT 2000), CIB, Reykjavik, Iceland. - available at <http://www.ds.arch.tue.nl/Research/Publications/Jos/CIT2000/CIT2000vanLeeuwen.pdf>
- Wang W., Zmeureanu R. and Rivard H. (2005)**, "Applying multi-objective genetic algorithms in green building design optimization", Building and Environment, Vol. 40, pp. 1512-1525. - available at <http://www.sciencedirect.com/science/article/pii/S0360132304003439>
- Watson I. (1997)**, Applying Case-Base Reasoning: Techniques for Enterprise Systems, Morgan Kaufmann, San Francisco, CA.
- Yan-chuen L., Phil M. and Gilleard J.D. (2000)**, "Refurbishment of building services engineering systems under a collaborative design environment", Automation in Construction, Vol. 9 No. 2, pp.185-196. - available at <http://www.sciencedirect.com/science/article/pii/S0926580599000047>
- Yu A.T.W., Shen Q., Kelly J. and Hunter K. (2005)**, "Application of value management in project briefing", Facilities, Vol. 23 No. 7/8, pp.330-341. - available at <http://www.emeraldinsight.com/journals.htm?articleid=1502620&show=abstract>
- Yusuf F. and Alshawi (1999)**, "Improving the brief through information and process modeling", in Lacasse, M.A. and Vanier, D.J. (Eds.), Proceedings of the CIB W78 8th International Conference on Durability of Building Materials and Components, Institute for Research in Construction, Ottawa, Canada, pp. 2651-2663. - available at <http://www.irbnet.de/daten/iconda/CIB2250.pdf>
- Zavadskas E. K., Kaklauskas A., Vainiunas P., Dubakiene R., Gulbinas A., Krutinis M., Cyras P. and Rimkus L. (2006)**, "A building's refurbishment knowledge and device based decision support system", in Luo, Y. (Ed.), Cooperative Design, Visualization, and Engineering - LNCS 4101, Springer, Berlin, pp.287-294. - available at <http://www.springerlink.com/index/3827868668881158.pdf>.