L. Rocha

Master in Informatics and Computing Engineering
Inferência Situacional em Sistemas Baseados em Informação de Contexto
Luís Miguel Marques Pereira Rocha
—

Abstract

Since the early 90’s, many entities started to collect data about their users in order to extract new information that might be usefull to improve their business and client relationship. The goal is to provide products and services that take into account the context and preferences of its users at a given moment. Tecnology is evolving in such a way that in a near future anyone may be permanently connected to the information networks. The recent mobile device and wireless networks expansion provided many new possibilites to gather and distribute this kind of data, making it possible to collect a massive number of context data in much less time.

PT inovação wants to explore this concept, creating a business model based in Context-asa- Service (CXaaS). In order to do this there is a framework currently under development that allows to specify a collection of libraries to assist the comunication between diferent applications in a context-based system. However, most of the retrived data is not significant enough for the purpose of creating context-oriented services.

The main goal of this work included the development of another kind of provider, in which the system could centralize most of the data processing in order to deduce new high-level information from the low-level information that is being collected. In order to accomplish this goal, a studywas done in the area of existent context management systems and different approaches for context representation and inference were analysed. Also, the most common geografic clustering algorithms and location classification methods were studied.

The developed component was mostly implemented using available Java technologies and is composed by 4 modules: a central component, a machine learning framework, a rule engine based in Jess and an ORM (Object-Relational Mapping) interface with a MySQL relational database. The system makes it possible to integrate any kind of context information that may be specified in an easy and quick way and also uses a set of defined rules to produce high-level information from the low-level collected data.

This system was tested for gathering GNSS (Global Navigation Satellite Systems) information from an Android application and process that data afterwards in order to find which are the most significant locations in the user daily routine. These locations are later classified according to a specified set of categories, as well as in terms of the importance that these places may have to each person.

The evaluation done to the system shows that it is possible to implement an incremental clustering approach over the gathered data in order to efficiently identify the most relevant locations in the daily routine of the users. This apprach also grants a system response under a very acceptable time. The classification methods based on Google Places API queries, however, did not show reliable results for the proposed goals.

Resumo

Desde o início da década de 90 que diversas entidades começaram a recolher de informação sobre os utilizadores para posteriormente extrair informação relevante que lhes permita melhorar o seu negócio e a sua relação com os seus utilizadores. Começou a ser, assim, possível fornecer produtos e serviços que têm em conta o contexto e preferências dos seus utilizadores a cada momento. As tecnologias estão a evoluir cada vez

mais no sentido de, num futuro próximo, qualquer pessoa estar quase permanentemente ligada às redes de informação. A recente expansão dos dispositivos móveis e redes sem fios veio potenciar esta realidade, sendo hoje muito mais fácil recolher uma enorme quantidade de dados sobre o contexto de cada pessoa.

A PT Inovação pretende explorar estes conceitos e desenvolver um modelo de negócios baseado em context-as-a-service (CXaaS). Para este efeito, já se encontra a desenvolver uma plataforma que permite definir bibliotecas que facilitam a comunicação entre diversas aplicações de um sistema, nomeadamente o envio da diversa informação que é recolhida sobre os seus utilizadores. No entanto, a informação que chega da maioria destas aplicações não é suficientemente significativa para prestar serviços dirigidos ao contexto em que cada pessoa se encontra.

O objectivo do trabalho aqui descrito passou por desenvolver um novo tipo de componente no qual se pudesse centrar o processamento da informação de mais baixo nível que é actualmente recolhida, de forma a inferir nova informação de mais alto nível. Foram estudados alguns sistemas de gestão de contexto existentes e os tipos de abordagem usados na representação e inferência sobre a informação do sistema. Foram também estudadas as principais abordagens de agregação de posições geográficas e métodos de classificação de locais.

O componente desenvolvido foi implementado recorrendo a diversas tecnologias baseadas em Java e é constituídos por 4 módulos: um componente central, um módulo de aprendizagem computacional, um motor de regras baseado em Jess e uma interface ORM (Object-Relational Mapping) com uma base de dados relacional MySQL. O sistema possibilita ainda uma fácil integração de qualquer tipo de informação de contexto, tanto sobre os tipos actualmente definidos como qualquer outro que possa vir a ser definido no futuro. Possibilita também a inferência de informação de mais alto nível através de um conjunto de regras que podem ser definidas à medida das soluções pretendidas.

Durante o período em que decorreu este trabalho, o sistema desenvolvido foi testado de forma a recolher posições GNSS (Global Navigation Satellite Systems) obtidas a partir de uma aplicação desenvolvida para dispositivos com plataforma Android e posteriormente processar esses dados de forma a determinar quais o locais mais frequentes no dia-a-dia de cada utilizador. Estes locais são classificados segundo um conjunto de categorias prédefinidas assim como em termos da importância relativa para o utilizador.

As avaliações efectuadas sobre o sistema mostram que uma abordagem de agregação incremental sobre os dados permite ao sistema ir identificando eficientemente os diversos locais à medida que novos dados vão sendo recolhidos. Esta abordagem garante também um bom tempo de resposta por parte do novo componente desenvolvido. Os métodos de classificação baseados na API do Google Places, no entanto, demonstram não ser minimamente fiáveis para os objectivos propostos.

Jury

- Chair: Luís Filipe Pinto de Almeida Teixeira
- External Examiner: Maria Benedita Campos Neves Malheiro
- Supervisor: João Correia Lopes
- Date: 13/7/2011

From:
https://web.fe.up.pt/~jlopes/ - **JCL**

Permanent link:
https://web.fe.up.pt/~jlopes/doku.php/students/201107l-rocha

Last update: **12/10/2012 19:00**