Goals in Social Media, Information Retrieval and Intelligent Agents

Dimitra Papadimitriou [1*], Yannis Velegrakis [2*], Georgia Koutrika [3**], John Mylopoulos [4*]

*University of Trento, Italy **HP Labs, Palo Alto, USA

[1] papadimitriou@disi.unitn.it [2] velgias@disi.unitn.eu [3] koutrika@hp.com [4] jm@disi.unitn.it

***Abstract*—This tutorial provides a comprehensive and cohesive overview of goal modeling and recognition approaches by the Information Retrieval, the Artificial Intelligence and the Social Media communities. We will examine how these fields restrict the domain of study and how they capture notions easily perceived by humans' intuition but difficult to be formally defined and handled algorithmically. It is the purpose of this tutorial to provide a solid framework for placing existing work into perspective and highlight critical open challenges that will act as a springboard for researchers and practitioners in database systems, social data, and the Web, as well as developers of web-based, database-driven, and social applications, to work towards more user-centric systems and applications.**

I. Introduction

Human activity is almost always goal-oriented, be it in a physical context or as part of an interaction with a computer system. Goals are ubiquitous in human and social activities, explaining, rationalizing, providing context and meaning for actions that human, social or artificial actors perform. Goals explain things people do in their personal life e.g., "learn Portuguese", "watch the Bolshoi ballet", "find a job". They also account for business and professional activities, e.g., "finish project B" and "increase sales by X%". By their very nature, they can be fulfilled in multiple ways that define a space of alternative sequences of actions that one can choose to perform. Goal-oriented behavior makes human activity purposeful and intentional, as opposed to aimless and random.

In computer science, goals may be present in almost every step of the *life cycle of a system or service*, from the analysis and design to the development and actual operation of the system, as well as in its maintenance phase. Goals may offer us a new approach for modeling and profiling users, as well as incorporating common sense knowledge into systems. In the phase of requirements analysis, we meet goal-oriented approaches that are contrasted with approaches concentrated on the functionality of the system-to-be, i.e., on what the system is supposed to do [1]. By taking into consideration goals into the development of a system, we can resolve the trade-off between performance, flexibility, cost and security, and furthermore enhance systems and services with features of adaptivity and personalization [1], [2]. If we manage to infer a user's goal, we can guide her directly to the part of the system that is "responsible" for the specific goal or adapt the underlying algorithms of the application respectively to the user goal without irrelevant intermediate steps or provision of additional feedback and unnecessary computations that raise the cost and reduce the performance of the system [3], [4], [5], [6].

During system operation, knowledge of user goals can help a system offer a significantly improved and more engaging experience. For example, it can lead to customization (at the time a system is deployed) and personalization (at the time a user first engages with a given system) or facilitation (as the user is trying to fulfill one of her goals). We can also prevent the user from misusing or causing damage to the system deliberately or not [7]. Moreover, knowledge about goals may be "transferred" from social networking sites, blogs or other web applications to another system so as to offer additional functionality or services that require knowledge about user goals and user actions towards satisfying these goals [8], [9].

Goals in computer science have been used in environments phenomenally disconnected, from physical locations monitored by sensors where users act, e.g., an airport monitored for suspicious behavior, to computer software or software for electronic devices such as CD/DVD devices, the Web as a collection of resources accessible through query answering, and the real world modeled through the analysis of social data, e.g., how and why people learn a foreign language. Goals have been studied in disparate fields, from artificial intelligence to information retrieval and HCI, and they have been used in the development of intelligent systems, web information retrieval systems, and social applications. As a result, we witness a seemingly disconnected mosaic of methodologies regarding how goals are perceived, modeled, and implemented. From a terminology standpoint, sibling concepts such as intentions, actions, and motivations have been used along with goals further complicating the landscape.

This tutorial serves the following objectives. We will present how goals are perceived, modeled, and implemented by the Information Retrieval, the Artificial Intelligence and the Social Media communities. We will examine how these fields restrict the domain of study and how they capture notions easily perceived by humans' intuition but difficult to be formally defined and handled algorithmically. Our goal is to provide an interdisciplinary presentation that smooths out and aligns terminology as well as methodologies enabling.

This tutorial also offers technical background for researchers and practitioners with a general computer-science and/or mathematical and technological background that would like to develop more goal-oriented methods and systems. It is

[1*]The author has been supported by the ERC Advanced Investigation Grant N.267856"

also the purpose of this tutorial to provide a solid framework for placing existing work into perspective and highlight critical open challenges that will act as a springboard for researchers and practitioners in database systems, social data, and the Web, as well as developers of web-based, database-driven, and social applications, to work towards more user-centric systems and applications. We discuss how goals and goal-oriented methods can get infused into research in database systems as well as web and social applications and outline research challenges and opportunities.

II. Outline

Our tutorial is structured in the following parts.

A. Goal Framework

We will present a framework that formally defines goals and the interrelated terms (such as motivations, actions, and intentions) and tasks (such as plan and goal recognition). In the subsequent parts of the tutorial, we use this framework as a common way to describe, align, and compare the various approaches. We will also show that the wide range of applications and services that are meant for use in phenomenally disconnected environments and often approached from different perspectives can be examined through the same prism.

B. Goal-oriented Systems and Applications

We will present several systems where goals play a central role from story generators and intelligent interface agents to web search engines and social networking sites. For these applications, we will illustrate several aspects of their goal-oriented behavior, including: (a) what is considered as the environment under study, (b) how goals, intentions or motivations, and actions are modeled, and (c) which type of data is used and how they are gathered. We will also present particular features and challenges in each category of systems.

For example, intelligent agents are computer-based software programs that make decisions that serve better long-term goals, i.e., they exhibit proactive or otherwise goal-driven behavior. Intelligent agents have been mainly studied by the Artificial Intelligence community. The goal-driven nature of intelligent agents requires deep knowledge of the problem domain that should be correctly decoded in the used algorithms in order to recognize at any time the goal that the agent follows as well as the actions and/or environmental conditions that may potentially lead to the fulfillment or failure of the goal. On the other hand, in web search, user queries are matched against a specific goal in a pre-existing taxonomy and the goal inference should take place right after the first query or in the worst case after a couple of queries. Once matched, the characteristics of the resources of the respective taxonomy are used in order to select among all the possible query results the resources that have similar characteristics to those of the selected goal from the taxonomy.

The Goal Framework and Goal-oriented Systems and Applications sections comprise the first part of the tutorial.

C. Goal Modeling and Recognition

To introduce goals in some application scenario, the system needs to perform two principal tasks. The *first* is the collection and modeling of the required knowledge about goals and the *second* is the recognition of the goals and the plans by observing the actions of an actor using that knowledge and the reaction of the system according to that. Goals and intentions cannot be easily captured. Analyzing user actions such as clicks and purchases can reveal patterns and behaviors but understanding the goals behind these actions is a different and challenging issue. We will present several approaches from AI, web information retrieval, and social data analytics on goal modeling and recognition.

1) Consistency approaches: By consistency we refer to the approaches in which the model should be sufficient to match any set of observations to a latent goal and subsequently to a plan of one of the goals which are possible to be implemented within the examined environment. The matching task does not always lead to some result but when it does it is always one and only one goal.

In this category, we will present: *plan libraries*, *consistency graphs*, and *action-centric representations*. Plan libraries are the outcome of the effort of a number of experts in a domain to encode their knowledge in a way that enables automatic goal recognition. They could be characterized as a set of recipes describing alternative plans for implementing a set of goals in which the developers of the goal recognition system are interested. Instead of constructing a complete plan library that includes all the possible plans that are related to every possible goal, consistency graphs focus on defining what constitutes a valid plan. In other words, how the allowable actions can be combined to a plan that fulfills a goal according to the structure, the restrictions, and the functionality of the system under study. Action-centric representations originally proposed for classical planning problems have recently been used to model goal and plan recognition by exploiting the progress in modern plan synthesis.

2) Taxonomy-based approaches: Taxonomy-based approaches capture user goals in a taxonomy by analyzing user actions in the specific application or domain in combination with expert knowledge. Such approaches have been typically adopted in web search as well as sponsored search.

We will present representative taxonomies for web search and sponsored search. The various goals behind user queries are identified by analyzing previously executed queries and the results that were selected to best represent the answer to the query. A number of taxonomies were derived from extended user studies using questionnaires and interactive tools on web browsers tracking user moves such as clicks and form submissions.

Based on the taxonomy, a model on the characteristics of the resources that best serve a particular goal can be identified. We will describe representative approaches that have used *rule-based annotators* as well as *automatic classifiers* such as Support Vector Machines (SVMs) and decision trees for goal recognition.

3) Corpus-based/Statistical approaches: Corpus-based methods, are based on a plan corpus which consists of a set

of alternative plans for a set of goals. The plan corpus is used as a training set for statistical models that can make predictions for future observations. Hence, in contrast to taxonomy-based and consistency approaches, in corpus-based approaches there is no ground truth about the environment; uncertainty expressed in probabilities is the factor which rules the outcome of the recognition process.

Despite the difficulty in obtaining a satisfactory plan corpus, there has been a lot of work on statistical approaches for building probabilistic models to predict goals within systems with artificial intelligence. The two most widely used classes of probabilistic graphical models are: *Markov models* and *Bayesian networks*. The main task of Markov models used for goal recognition is to compute the conditional probability of a sequence of observations given some evidence i.e., to check to which extent the current observations would be justified. We will provide an overview of several variations of Markov Models. Bayesian networks have been widely used for goal recognition in artificial intelligence because they manage to capture causality among actions and goals.

4) Real-world approaches: In all the approaches above, the set of goals and/or plans are a-priori known, well-defined and limited to a certain domain. In real-life, however, there is an infinite set of goals and an infinite world of possible actions and plans to achieve these goals. Social media could be considered as "a snapshot of the real world"; with its own special rules though. User goals may regard users "life" within the social platform, e.g., a user goal to increase her number of friends by 100, or real-life. The first type of goals can be fulfilled by actions within the platform such as likes, posts, or follows, while the second type is fulfilled by real-life actions such as run, buy, and read. In this section of our goal modeling and recognition part of our tutorial, we will examine how real-life goals can be recognized and leveraged from social media by analyzing user-generated content, mostly textual descriptions, in social media. Textual descriptions usually can be interpreted in multiple ways. Consequently, by performing text analysis it may not be feasible to extract hard goals but soft goals that allow uncertainties.

D. Research Opportunities and Challenges

We will conclude our tutorial with a discussion of how goals and goal-oriented methods can get infused into research in database systems as well as web and social applications and outline research challenges and opportunities.

REFERENCES

[1] J. Mylopoulos, L. Chung, and E. Yu, "From object-oriented to goal-oriented requirements analysis," *Commun. ACM*, vol. 42, no. 1, pp. 31–37, 1999.

[2] A. Broder, "A taxonomy of web search," *SIGIR Forum*, vol. 36, no. 2, pp. 3–10, 2002.

[3] M. Armentano and A. Amandi, "Goal recognition with variable-order markov models," in *IJCAI*, 2009, pp. 1635–1640.

[4] S. Carberry, "Tracking user goals in an information-seeking environment," in *AAAI*, 1983, pp. 59–63.

[5] K. Gold, "Training goal recognition online from low-level inputs in an action-adventure game," in *AIIDE*, 2010.

[6] S. Zhe, T. Xia, and X. Cheng, "Modeling users' information goal transitions and satisfaction judgment: Understanding the full search process," in *Web Intelligence*, 2010, pp. 431–434.

[7] P. A. Jarvis, T. F. Lunt, and K. L. Myers, "Identifying terrorist activity with ai plan recognition technology," in *In Innovative Applications of Artificial Intelligence (IAAI)*. AAAI Press, 2004, pp. 858–863.

[8] M. Castellanos, M. Hsu, U. Dayal, R. Ghosh, M. Dekhil, C. C. Limon, M. Puchi, and P. Ruiz, "Intention insider: discovering people's intentions in the social channel," in *EDBT*, 2012, pp. 614–617.

[9] S. Louvigne, N. Rubens, F. Anma, and T. Okamoto, "Utilizing social media for goal setting based on observational learning," in *Advanced Learning Technologies (ICALT)*, 2012, pp. 736–737.

BIOGRAPHIES

Dimitra Papadimitriou is a PhD student and a member of the dbTrento (Data and Information Management) group at the University of Trento, Italy. She received her MSc and BSc degree in Computer Science from Aristotle University, Greece. She has also spent time as a visitor at HP Labs, Palo Alto, California. Her scientific interests include large scale data mining and information extraction with a focus on user-generated data.

Yannis Velegrakis is a faculty member of the Department of Information Engineering and Computer Science of the University of Trento. He holds a PhD degree in Computer Science from the University of Toronto. His research areas of expertise is Big Data Analytics, Social Data, and Large Scale Information Management. Prior to joining the University of Trento, he held a researcher position at AT&T Research Labs, and has been a visitor at the University of California, Santa-Cruz, the IBM Almaden Research Center, and the Center of Advanced Studies of the IBM Toronto Lab. He has served in many program committees of Data Management Conferences and has been the general chair for VLDB 2013, WebDB 2012, DESWEB 2010/11 and SWAE 2007. He holds 2 US patents and has been a Marie Curie Fellow for the period 2006-2008.

Georgia Koutrika is a senior research scientist at HP Labs, in Palo Alto, California. In the past, she was with the Information Integration Group at IBM Almaden Research Center, and a postdoctoral researcher at Stanford University. Her research focus is on inventing and developing innovative technologies around user analytics and profiling, personalization, and smart content selection, with applications including search, automated publishing, education, and marketing. Georgia has co-authored more than 60 papers in peer-reviewed conferences (including ACM SIGMOD and IEEE ICDE) and journals (such as ACM TODS and IEEE TKDE). She is general co-chair of SIGMOD 2016 and she has served as the general co-chair/PC chair for several workshops and as PC member for SIGMOD, ICDE, WWW, and KDD.

John Mylopoulos holds a professor emeritus position at the University of Trento (Italy) and University of Toronto (Canada). His research interests include conceptual modeling, requirements engineering, data semantics and knowledge management. Mylopoulos is a fellow of the Association for the Advancement of Artificial Intelligence (AAAI) and the Royal Society of Canada (Academy of Sciences). He has served as program/general chair of international conferences, including IJCAI (1991), Requirements Engineering (1997, 2011), and VLDB (2004). He was recently awarded an advanced research grant from the European Research Council (ERC) for a project titled "Lucretius: Foundations for Software Evolution".