"Context-Oriented Software Adaptation – A Brief State of the Art"

Mens, Kim ; Cleve, Anthony ; Dumas, Bruno

ABSTRACT

In this document we provide an extensive, though not exhaustive, overview of existing related research in the areas of behavioural adaptation, user interface adaptation and database adaptation.

CITE THIS VERSION

Abstract. In this document we provide an extensive, though not exhaustive, overview of existing related research in the areas of behavioural adaptation, user interface adaptation and database adaptation.

1. Behavioural Adaptation

This section provides an overview of the research that has been conducted on approaches that support the development of software systems allowing a consistent and dynamic adaptation of their behaviour according to their current internal state and the situation of their surrounding environment. System monitors and sensors provide information about the various situations to which software systems could adapt. Dynamic software adaptation of the system takes place with the objective of providing a more appropriate behaviour and quality of service of the software system with respect to the current situation of its execution environment. This requires such software systems to:

1. provide the necessary mechanisms for continuous adaptation, allowing components to be introduced into or removed from the system at any time, without stopping the system’s execution;
2. ensure that the system’s adaptations remain consistent with the expected behaviour of the system and other adaptations made to the system.

These characteristics are important to ensure that the system effectively changes to provide a more appropriate service with respect to its changing environment, and to ensure that it behaves as expected under all possible situations to which it may be adapted and all combinations of available adaptations. Different mechanisms have been proposed to enable dynamic software adaptation with respect to the environment. Ensuring consistency of these adaptations, however, is challenging, given that the number of possible adaptation combinations increases exponentially with the introduction of every new adaptation.

Regarding the mechanisms to achieve dynamic software adaptation, following Cardozo [BH1] we distinguish three categories, according to the abstraction level at which the particular mechanisms for achieving behavioural adaptation are implemented: architectural solutions, middleware solutions and language solutions. Alternative but equally interesting taxonomies that focus more on the composition mechanisms or programming technologies are given, respectively, by McKinley [BH2] and Bainomugsha [BH3].

Architectural Solutions

The category of architectural solutions rely essentially on using dedicated software architectures and designs to enable adaptation of behavioural adaptations at run time.

Design patterns [BH4, BH5], for example, provide recurring solutions to particular design problems, often related to reusability and maintainability. Particular design patterns, such as the state pattern, strategy pattern, decorator, abstract factories and dynamic proxies, provide dedicated solutions for enabling dynamic system adaptations. Although these patterns do enable behavioural adaptations of a system, the solutions are rigid in the sense that the adaptations need to be explicitly foreseen and the system needs to be structured in a particular way to accommodate these adaptations.

Dynamic software updates [BH6, BH7] address the need to upgrade the behaviour of software systems without stopping them, while respecting consistency and robustness of the system, for example when adding new features, bug fixes or patches to a software system. Most approaches for dynamic software upgrades support only forward evolution of adaptations however, that is, adaptations can be introduced but not removed from the system after they have been introduced.

Software product lines [BH8] are a structured approach to design not just a single but an entire product line of different software products. They support customization of the product line into a single product either statically, at design time, by selecting the appropriate components and features building up a product, or even dynamically by generating or transforming the required system components at run-time.
Such **dynamic software product lines** [BH19, BH110] rely on different techniques to enable such transformations, for example, regenerative, composable components construction, or incremental-move [BH18]. Using dynamic software product lines effectively allows the reconfiguration of systems at run time. However, again all system adaptations need to be foreseen in order to create the variability model describing the adaptations and to specify the concrete transformation strategy. Moreover, the points of the program in which adaptations should take place must also be defined beforehand.

The main drawback of these architectural solutions is thus that all adaptations need to be anticipated beforehand, which may be too strong a constraint for the setting of highly changing environments.

**Middleware Solutions**

The category of **middleware solutions** explores platforms and frameworks focused on reducing the complexity of building software systems that can dynamically adapt their behaviour, by shifting complexity from the application design towards a reusable middleware architecture. Most of the existing middleware approaches that deal with adaptation of a system at run time are targeted to work in a distributed environment. Since in this particular research project we are less interested in the technical communication issues between different remote components of a system, in this section we only discuss those aspects of middleware solutions related to dynamic adaptation of a system.

**Dependency injection** [BH11], sometimes referred to as Inversion of Control, achieves dynamic adaptation at the granularity of software components, allowing components to be replaced freely by other components, as long as they respect a common interface. Different components can be dynamically injected into the application according to the current situation. But as for the architectural solutions, such behavioural adaptations must be anticipated by the programmer, and the system components need to be designed with this kind of adaptation in mind. This approach is thus less suited for the definition of fine-grained behavioural adaptations, and provides little support for managing the subtle interactions that may occur between two independently injected components or to combine different behavioural adaptations for a same component.

**Service-oriented architecture** (SOA) [BH12] is a middleware approach for the construction of a system from loosely coupled services. Each component of behavioural functionality is designed and provided as an independent service of the system. Inter-service communication takes place through standard interfaces, usually using XML or a Web Services Description Language (WSDL). A limitation of SOA is that adaptations are defined at a coarse-grained level, allowing only complete services to be adapted. In recent years different approaches have been proposed to allow a dynamic interchange of services in SOA, such as context-aware mashups [BH13].

**Cloud computing** [BH15] provides an infrastructure where storage and computation services are leased to users and adapted dynamically to particular situations such as user load and available computation resources. It has opened the door to the new trend of **Software-as-a-Service** (SaaS) [BH14] where software products are provided to the end user as a combination of different accessible services. Combining these ideas also rise to our vision of **applications as feature clouds** [BH16], providing software services that are composed dynamically from a set of available fine-grained features. Such a feature cloud programming model can be realized by relying on context-oriented programming technology (see further), enabling new or adapted features to be added to or removed from running applications on demand, according to contextual information, and taking into account feature dependencies.

**Event systems** [BH17] provide a middleware platform to program systems in terms of actions that are executed automatically whenever certain conditions are satisfied. This makes them an interesting candidate to build systems supporting dynamic adaptations. However, they present some drawbacks for defining adaptations at different levels of granularity, and for defining adaptation events independently of the core functionality of the system.

**Self-adaptive systems** [BH18], also known as autonomic computing or self-managing systems, are conceived as closed-loop systems with feedback from their self and from the surrounding context. Müller et al. [BH19] advise highly dynamic software systems whose operations may need to adapt dynamically to changing requirements and uncertainty in their execution environments, to be instrumented with self-adaptation mechanisms to monitor selected requirements and environment conditions to assess the need for evolution, plan desired changes, as well as validate and verify the resulting system.
Language Solutions

The category of language solutions explore programming paradigms that provide language facilities to enable the definition and introduction of adaptive behaviour in software systems.

Metaprogramming facilities [BH20] extend existing programming languages with the ability to write programs that can reason about and manipulate themselves. As such, they can be used to write programs that can dynamically modify their own behaviour. However, since they don’t provide any dedicated higher-level abstractions to represent contexts and behavioural adaptations to contexts, they are mostly useful as an underlying technique based on which more high-level language solutions can be developed.

Aspect-oriented programming [BH21] tackles the problem of modularizing behaviour that cuts across the base decomposition of a software system. By providing a notion of aspects that can be woven inside the base code of the system (either statically or dynamically) it can be used to define and apply behavioural adaptations to a system’s base behaviour, but still remains mostly a technical solution.

Context-oriented programming languages [BH22] take this idea of separation of concerns between the base behaviour of a system and its behavioural adaptations further, by providing dedicated language constructs for representing and managing contexts, context dependencies, and behavioural adaptations to context. As such they provide a high level of dynamicity for the introduction and withdrawal of fine-grained behavioural adaptations. Many COP languages have been proposed [BH22, BH23, BH24, BH25, BH26, BH27, BH28, BH29, BH30, BH31, BH32, BH33, BH34, BH35, BH36, BH37, BH38], most of which have been implemented as extensions to traditional object-oriented languages by relying upon metaprogramming or aspect-oriented programming techniques.

Feature-oriented programming [BH39, BH40] is akin to context-oriented programming [BH41, BH42] but grew from the software product line and feature modelling research communities, as a mechanism to write programs in terms of software features and their interactions, and even to activate and deactivate such features at runtime. We have recently shown that both paradigms show quite some similarities [BH42] and have started to explore a new generation of context-oriented programming languages by taking up ideas from feature-oriented programming [BH16]. Although many other programming paradigms exist that provide mechanisms for defining dynamic behaviour adaptations, such as reactive programming [BH3, BH43, BH44] and agent-oriented programming [BH45, BH46], because of the built-in abstractions it provides for dealing with context and adaptation to context, the paradigm of context-oriented programming seems to be one of the most promising ones to truly achieve context-oriented adaptation. Nevertheless, useful ideas regarding how and when to adapt and how to manage the interaction between different behavioural adaptations, can be borrowed from these other paradigms.

Partly orthogonal to the above classification of mechanisms to achieve dynamic behavioural adaptation, the class of dynamically adaptive software systems could also be classified according to the kinds of techniques, models and formalisms they provide for managing the consistency of such systems. In his PhD thesis on the identification and management of inconsistencies in dynamically adaptive software systems [BH1], conducted under the guidance of Prof. Kim Mens at UCL, Nicolás Cardozo conducted an exhaustive survey of models for conflict resolution and inconsistency management. He distinguished four different kinds of approaches. A first kind are the architectural approaches that use models to guarantee consistency of the system not only at design and development time, but also during its execution or evolution [BH17, BH48], to monitor and verify particular aspects of the system's behaviour. A second kind are the formal approaches, which rely on algebras (e.g., process algebras or coalgebras), logic (e.g., modal or temporal logic), or model checking (e.g., SAT solving), to prove or verify interesting properties of the system. A third kind are the rule-based approaches that rely on logic programming or other rule-based engines to manage and coordinate the behaviour of dynamically adaptive software systems. The fourth kind he distinguished are the state-machine based approaches that use automata and labelled transition systems, statecharts, dataflow graphs or Petri nets to model, verify and manage the expected states and actions of a dynamically adaptive software system. In particular, the formalism of Context Petri nets [BH1, BH49, BH50] enables the definition of adaptations and their allowed and expected interactions in a context-oriented software system, and provides certain consistency guarantees about the system at design-time, while also serving as a runtime mechanism to ensure that no inconsistencies are introduced. Other authors such as Filieri et al. [BH51] have explored probabilistic approaches to verify or guarantee the consistency of self-adaptive systems at run-time.
2. Adaptation of User Interaction

User interface adaptation is the process of modifying a user interface based on a number of parameters [AU11, AU12, AU13]. The interface can be adapted either when creating the interface, or dynamically in real time. The input as well as the output channels of the interface can be adapted. Parameters used when defining how to adapt the interface typically include three aspects: the user profile, the device profile, and environmental parameters [AU14], with fewer works also exploring the social surroundings of the user [AU15, AU16]. For example, an application could adapt displayed information based on the current activity of the user: if the user is in a car, the information presented could be related to current traffic jams while if the user is walking in a town, showing information on public transportation available in the immediate surroundings may be more appropriate.

Frameworks in User Interface Adaptation

When considering user adaptation, different ways of adapting the user interface can be considered. On the one hand, the user interface can be created in such a way that it adapts automatically. On the other hand, solutions exist where users are given the possibility to modify the user interface in a proactive way. Malinowski et al. [AU17] proposed a complete taxonomy to describe user interface adaptation. Their taxonomy is based on four different stages of adaptation in a given interface: initiative, proposal, decision and execution. Each of these adaptations can be performed by the user, the machine or both. Researchers have also explored how frameworks could help the definition of adaptive interfaces [AU18]. López-Jaquero et al. [AU19] extend Malinowski’s taxonomy into the ISATINE framework with seven stages of adaptation: the goals for user interface adaptation, the initiative for adaptation, the specification of adaptation, the application of adaptation, the transition with adaptation, the interpretation of adaptation as well as the evaluation of adaptation. Calvary et al. [AU14, AU10] presented a model-based user interface framework for multi-target interfaces (also called plasticity of user interfaces), allowing the description of design and runtime phases without taking into account specific implementation requirements. Multi-target refers to multiple contexts, where context denotes the context of use of an interactive system described in terms of a user, platform and environment model. The user model contains information about the current user(s), such as user preferences or limitations of disabled users. The platform model describes physical characteristics of the device the system is running on, including screen size or processor speed. Finally, the environment model contains information about social and physical attributes of the environment in which the interaction takes place. Beyond multi-target interfaces, multiple studies have explored how an interface can migrate from one device to the other [AU11, AU12] or even be distributed on multiple devices [AU13, AU14].

Adaptive Multimodal User Interfaces

While adaptation has been investigated mostly for traditional Window/Icon/Menu/Pointer (WIMP) interfaces [AU15] and context-aware interfaces [AU16], work devoted to the automatic adaptation of multimodal interfaces based on contextual information is more limited. Multimodal interfaces are user interface relying on multiple input and output modalities (such as speech, gesture, emotions…) to enable the communication between a user and her device [AU17]. Since early 2000, researchers have been exploring the possibilities of linking multimodal interfaces and user interaction adaptation [AU18]. Evidence that context influences a user’s preferences while interacting with a mobile device has been found by Lemmelä et al. [AU19]. Additionally, they identified which modalities and combinations of modalities best suit different situation requirements. For test purposes, an SMS application was developed and evaluated in a car as well as in a pedestrian context. Speech input was assigned as the default interaction technique for the car environment whereas 2D gestures (e.g. finger strokes) and motion gestures (e.g. tilt) were used in the walking environment. Doyle et al. [AU120] conducted a review and analysis of existing map-based multimodal systems. They further proposed and evaluated a novel multimodal mobile geographic information system (GIS) called Compass. Parameters such as effectiveness and efficiency of unimodal interaction were evaluated. To deal with the complexity of working with digital maps on small screen displays, Ramsay et al. [AU21] proposed the use of motion gestures like tilting an external device backwards and forwards to navigate within a map. Reufer et al. [AU122] investigated different motion gesture actions by using built-in smartphone sensors. They proposed a multimodal game using speech, 2D gestures and motion gestures as input modalities with
different actions specifically assigned to each modality. For example, 2D gestures like touch on the display were used for selecting and erasing content whereas motion gestures such as rolling (rotating the device around the z-axis) or pitching (rotating the device around the x-axis) were used to control different game options. In Bühler et al. [AUI23] the major concern was the flexible control of interaction modalities so that users as well as the system were able to make transitions between the different modalities. They presented a conceptual framework that described five combinations of input and output modalities based on a user’s level of attention within a car and a pedestrian setting. Last but not least, David et al. [AUI24] proposed a mobile middleware to facilitate the development and maintenance of adaptive mobile applications. An instant messenger prototype was built to illustrate their approach. The application allowed users to read and write SMS messages using either the device’s keyboard or speech. The application adapted its user interface depending on the user’s movements. These last two works are particularly interesting, as they focus on the adaptation of multimodal input. Based on the results of these different researches, in Dumas et al. [AUI25] we further explored the possibilities of automatic adaptation of multimodal interfaces with a calendar application accepting accelerometer-based movements, gestures, RFID-tagged objects and multi-touch as input modalities.

3. Database Adaptation

In this section we study the state of the art in database adaptation, starting from database evolution in the large, via more fine-grained static database evolution mechanisms and runtime database adaptation mechanisms, to more recent research on context-aware databases.

Database Evolution in the Large

The general problem of database evolution is a largely explored problem within the data modeling, database engineering and data management communities. Database migration, database merging and database restructuring are popular database evolution scenarios that involve not only changing the data components of applications, but also rewriting some parts of the programs themselves, even when no functional change occurs. In general, such evolution patterns involve the joint modification of several mutually dependent system components, namely the data structures and constraints (i.e., the schemas), the mappings between the schemas, the data instances and the programs (i.e. the data manipulation behaviour). When the system evolves, the consistency that exists between these those artifacts must be preserved.

Database Evolution via Coupled and Bidirectional Transformations

A promising approach to formalize and support such a co-adaptation process is by means of coupled transformations (a.k.a. co-transformations). The general concept of co-transformation was defined by Lämmel as follows: “A co-transformation transforms mutually dependent software artifacts of different kinds simultaneously, while the transformation is centred around a grammar (or schema, API, or a similar structure) that is shared among the artifacts” [DB1]. Generally, ”two or more artifacts of potentially different types are involved, while transformation at one end necessitates reconciling transformations at other ends such that global consistency is reestablished” [DB2].

Several authors have investigated the use of coupled transformations in the domain of database evolution. The 2LT project [DB3] aims to formalize and to support such a particular instance of coupled transformations, namely two-level transformations, which involve a transformation on the level of types with transformations on the level of values and operations. This generic approach revealed to be applicable to the coupled transformation of database schemas, data instances, queries, and constraints [DB4].

The PRISM system [DB5, DB6] provides an integrated support to relational schema evolution that includes (1) a language for the concise specification of modification operators for relational schemas, (2) impact analysis techniques that allow to evaluate the impact of schema changes, (3) automatic data migration support, (4) optimized translation of old queries to work on the new schema and (5) full documentation of the changes involved by the evolution. Query adaptation derives from the schema modification operators and combines a technique called chase and back-chase for query rewriting and SQL view generation.
Tervilliger et al [DB7] propose a comprehensive approach to the co-evolution of conceptual schema, physical schema and mappings in the particular context of the evolution of systems based on Object-Relational-Mapping technology. Their main focus is on achieving some local consistency of the chosen mapping patterns over time.

In previous work [DB8], we proposed a co-transformational approach to data-intensive system evolution, by focusing on semantics-preserving database evolution scenarios such as platform migration and schema refactoring. According to this approach, each generic semantics-preserving schema transformation is associated to a set of abstract query transformation rules allowing to propagate successive database restructuring to the program level (thereby preserving global consistency). Those co-transformation rules can then be composed to support the successive application of schema refactorings. We show in [DB9] that the combination of transformational and generative techniques provides a sound basis for the co-evolution of database schemas and programs.

An interesting particular category of coupled transformations are bidirectional transformations. According to its most generally agreed definition [DB10], a bidirectional transformation (bx) between two sources of information A and B consists of a pair of unidirectional transformations: one from A to B and another from B back to A. In many cases, the flow of data from A to B dominates the flow of data from B to A. In these cases, A is called the input/source/master of the bx, and B is called the output/target/slave. The transformation from A to B is called the forward transformation, whereas the transformation from B to A is called the backward or reverse transformation.

Bidirectional transformations are often seen as a mechanism for preserving the consistency of two (or more) inter-dependent software artifacts over time. The use of bx has been actively investigated by researchers from many different areas including software engineering, database engineering, programming languages, and document engineering to solve diverse problems such as: (1) constructing updatable views in relational databases; (2) transforming, integrating, and exchanging data; (3) synchronizing replicas of data expressed in different formats; (4) computing and synchronizing views of software models; and (5) providing convenient interfaces for editing complicated data sources.

All the above-mentioned approaches supporting the coupled transformation of schemas, data and behaviour consider this co-adaptation problem at compile-time (not at runtime). In other words, the adaptation mechanisms require the system to be shut down before it can be adapted, recompiled and redeployed. Furthermore, to the best of our knowledge, the above-mentioned approaches to co-evolving databases and programs only consider unidirectional coupled transformations. In addition, most of those approaches are restricted to information-preserving database evolution scenarios, that involve semantics-preserving schema transformations.

**On-the-fly Database Adaptation**

Only a few authors have investigated methods, techniques and tools to support on-the-fly schema updates. Uddin et al [DB11] propose an approach allowing to apply schema changes at runtime, i.e., without the need to shut down the running data-intensive system. Idreos et al. [DB12] present a complementary adaptive indexing technique for optimizing data access performance at runtime without human intervention. Those runtime approaches mainly focus on schema and data adaptation; they do not propose solutions to support schema-behaviour co-adaptation at runtime.

**Context-Aware Databases**

Data-intensive systems manage increasingly complex and huge amounts of data that are suited for different kinds of users each performing tasks of different nature in different, changing contexts. Most of the effort for designing, maintaining and evolving these systems depends on their complex interactions with (big) data sources. From this perspective, another important problem that needs to be tackled is how to facilitate data-intensive system variability in terms of data and data access. This problem has recently started to be addressed in the literature of context-aware databases [DB13] by means of methodologies, techniques and tools for creating sub-portions of a global database based on different factors, such as current context, user tasks and user preferences [DB14, DB15]. Several approaches in the literature
consider external factors, i.e. the context, to provide the user with the most relevant data. Preliminary ideas concerning context-aware data tailoring have been introduced by Bolchini et al. in 2009 [DB13] along with a generic architecture to enable context-aware access to a data source.

Generic feature-based approaches have been proposed by the software engineering community with the aim of creating a target model starting from a source model. These approaches, which—as we will see shortly—have been a source of inspiration for the database engineering community, rely on either pruning (filtering) or merging techniques. For instance, in [DB16] the authors support a generic approach to generate UML product models by pruning the base model which implements all possible feature configurations. This method evaluates the presence conditions belonging to the fragments of the base model in order to assess which fragments to include in the product according to the selected features. This approach supports the consistency of the products with respect to the base model. Saval et al. [DB17] identify the main challenges and propose a set of good practices for implementing the feature-based merging approach whose aim is to create a target model by combining the model fragments belonging to the features to derive. The merging approach presented in [DB18] supports the product derivation of architectural models according to the target feature configuration.

In the research literature on context-aware databases, most of the approaches that provide context-based data variability also follow either the filtering or the merging perspective. Bolchini et al. [DB19] propose a merging method to extract the relevant portion of a relational schema based on a hierarchical context model [DB7]. Each context element corresponds to a different view of the global schema. By means of two operators, their method supports the creation of the view starting from basic excerpts of data belonging to active context elements. Villegas and Olive [DB21] define a filtering method to select an excerpt of the conceptual schema based on a set of entity types that are required in a certain context. The method creates a consistent excerpt of the conceptual schema by starting from an augmented set of entity types. The focus set is then incremented with the entity types that optimize a single utility function which combines the absolute importance of an entity type and its closeness to the focus set.

Ciaccia and Torlone [DB14] present an approach to select the most suitable data based on context-dependent user preferences. They propose a definition of context as a set of elements that are related hierarchically. Each context may have a different preference, each one expressed as a binary relation over the tuples of a database relation. The authors propose an operator to select the best tuples of a table according to a context-dependent preference relation.

Miele et al. [DB22, DB23] propose an approach to context-dependent data personalization, allowing the user to express preferences that specify which data (s)he is more/not interested in depending on each specific context. A partial order can be defined among the data, based on which only the top (most preferred) portion of the data are loaded in a given context. The complementary problem of modeling, composing and exploiting user preferences in the context of data management has been addressed by several authors in recent years. A survey of this domain is available in [DB24].

In recent work [DB25], we presented a theoretical framework for database schema adaptation relating context models, feature models and data models, by considering two levels of abstraction for the latter (conceptual and logical schemas). This feature-based, context-oriented adaptation framework includes, among others, three generic algorithms for adapting database schemas as a consequence of context variations. We also published preliminary yet promising results in the domain of run-time reconfigurations of context-oriented databases [DB26]. In particular, we identify a critical property of the system, i.e., stability as the capacity of the system to have the lowest possible variations of data w.r.t. variations of user-driven data requirements. This approach aims to optimize stability by exploiting a predictive model expressing future possible variations of informational needs. We have formalized, implemented and experimented a multi-objective decision-making process to determine the best reconfiguration of data depending on two relevant metrics representing the input data-oriented predictive model.
References to Research on Behavioural Adaptation


References to Research on User Interface Adaptation


References to Research on Database Adaptation


