Aggregating Operational Knowledge in Community Settings
(Short Paper)

Srinath Srinivasa

International Institute of Information Technology,
26/C, Electronics City, Bangalore, India 560100

Abstract. In a community setting, *operational knowledge* or "knowledge that works," refers to knowledge elements that can be *put to use*. This can be contrasted with *encyclopedic knowledge* or "knowledge that tells" that sets the background shared mental model among members of a population. Given any community, large amounts of operational knowledge are routinely diffused through social connections. While current day social media catalyzes such diffusion, they are not primarily suited to capture and represent knowledge. This paper argues that operational knowledge aggregation is in some sense, the opposite of encyclopedic knowledge aggregation. The latter is a *convergent* process that aggregates different local views into a global view; while the former is a *divergent* process, where common knowledge gets segregated into several local worlds of utilitarian knowledge. If the community as a whole is coherent, these different worlds end up denoting different *aspects* of the community's dynamics. Local worlds are not independent of one another and characteristics of one local world affect characteristics of other local worlds. To capture this phenomenon, a data model called Many Worlds on a Frame (MWF) is proposed, that is detailed in this paper.

1 Introduction

Operational knowledge or "knowledge that works" pertains to knowledge fragments that can be put to use. This can be contrasted with encyclopedic knowledge or "knowledge that informs" that sets the background mental model of a population.

This paper argues that the process of operational knowledge aggregation is in some sense, the opposite of encyclopedic knowledge aggregation. Encyclopedic knowledge *converges* from several local perspectives to one global model, while operational knowledge *diverges* from a set of commonly known knowledge elements, to different ways in which it can be put to use.

Operational knowledge management is characteristic of "communities" which are formed by humans having a *shared human condition* like – living in the same neighbourhood, having the same hobby or suffering from the same disease. This paper also contrasts the notion of communities, with the notion of *organization*

R. Meersman et al. (Eds.): OTM 2012, Part II, LNCS 7566, pp. 789–796, 2012.

Table 1. Contrasting between organization, community and crowd

	Organization	Community	Crowd
Structure	Highly structured	Loosely structured	Unstructured
Motivation	Collective vision, occupation, profession	Shared human condition	Herd instinct
Affiliation	Formal	Semi-formal	Informal or transient
Knowledge dynamics	Top-down	Bottom-up	Diffusion and cascades

and with *crowds*, to argue that community knowledge management is a unique problem in itself.

Finally, this paper addresses ontological frameworks and supporting services to capture the divergent dynamics of operational knowledge in a community. The result is a framework called Many Worlds on a Frame (MWF), augmented with a set of cognitive heuristics. The framework is partially implemented in a project called RootSet[1]. A more detailed version of this paper is available as a technical report [1].

2 Underlying Definitions

2.1 Organization, Community and Crowd

The term *organization* is used here to refer to groups that have formal, codified structures, collective vision and well-specified roles for members of the organization. Members are formally affiliated to an organization and the organization itself is treated as a separate abstract entity, distinct from its members.

At the other end of the continuum is the *crowd*, where this term is used to refer to loosely formed transient groups of people. They are not driven by any formal structure, and members are not even formally affiliated to the crowd. The formation of crowds can be explained by the latent *herding instinct* of humans. Humans naturally tend to congregate into herds, and herds often implicitly attract and exert influence on individuals and mould their behaviour [2,3].

The dynamics of crowds tend to be sporadic in nature and knowledge exchange happens by processes of *diffusion* and *information cascades*. Crowd dynamics also tend to be "stateless" without much memory of what happened earlier, since people's affiliation with the crowd are transient.

Communities are more structured than crowds, but much less structured than organizations. Communities are formed by people who share a common *human condition* like living in the same neighbourhood, suffering from the same disease, employed in the same profession, etc. Unlike a crowd, a community has a common

[1] Prototype RootSet implementation available from `http://rootset.iiitb.ac.in/`

denominator that unites all its members. However, unlike an organization, there need not be a formal structure and shared vision and goals for the community.

The motivation for people to join a community is to network with other "*kindred souls*" – people who share the same condition as them, so they can benefit from one another's experience. The dynamics of the community tend to be ad hoc, but not arbitrary. Communities have a much longer life than crowds and their dynamics do operate on memory of what happened in the past. Knowledge aggregation in communities tend to happen in a bottom-up fashion, driven by individual needs and experiences. This is in contrast to that of organizations where organizational vision is usually established top-down; and from crowds, where knowledge exchange are characterized by diffusion models.

Table 1 summarizes the differences between organization, crowd and community.

2.2 Encyclopedic versus Operational Knowledge

Encyclopedic Knowledge. Encyclopedic knowledge aims to inform and clarify underlying propositions that make up a subject area or topic. Its objective is to establish a common "global world-view" among members of a community concerning a given subject area or topic.

In other words, there are no *subjective versions* of encyclopedic knowledge, specific to individuals. Individuals may possess subjective *beliefs*, *perspectives* and *opinions*, but when a community seeks to codify its encyclopedic knowledge, individual perspectives are discussed and debated to arrive at a common representation. The quality of the aggregated encyclopedic knowledge is determined based on how well it balances different perspectives.

For instance, the core content policies in Wikipedia illustrate the objective nature of encyclopedic knowledge. They are based on the following three underlying principles[2]: *Neutral Point of View (NPOV)*, *Verifiability (V)* and *No Original Research (NOR)*, where NPOV requires that any piece of knowledge whose truth cannot be established, should be presented in a way that is bias-free.

Operational Knowledge. Operational knowledge in contrast, poses a different problem. These are utilitarian knowledge fragments that people in a community can *put to use*.

While encyclopedic knowledge aggregation flows from several local perspectives to one global world-view, operational knowledge aggregation flows from a commonly shared global world-view to several local perspectives. Encyclopedic knowledge *converges* from many local to one global, while operational knowledge aggregation *diverges* from one global to many local. Even though the process is divergent, we call it as "aggregation" because several local perspectives are aggregated by factoring knowledge elements from different parts of the common knowledge, and augmenting it with other knowledge elements to make them consistent and complete within themselves.

[2] Wikipedia Core content policies: `http://http://en.wikipedia.org/wiki/Wikipedia:CCPOL`

In cognitive science, convergent and divergent thinking styles are seen as the two important cornerstones of our thinking process [4]. A convergent process collates different inputs and produces one factual or "correct" answer. For example, a question like, "Is 17 a prime number?" has one factual answer. A divergent process on the other hand, starts with a fact and produces several "correct" answers. A question like "What are the different ways in which prime numbers can be put to use?" has several correct answers.

Given a community of people with a shared human condition that is common knowledge, the different ways they put their common knowledge to use, defines the different *aspects* of the community. Aspects represent consistent local "worlds" within a larger community. Aspects are internally self contained, but are not independent of one another.

Unlike encyclopedic knowledge, subjectivity in operational knowledge is very high. A piece of knowledge may be utilitarian or not, based on who is consuming it and what kind of use they have in mind. The subjective nature of operational knowledge means that users and privileges are an *integral part of knowledge representation*, rather than an element of the knowledge management application.

A divergent process means that knowledge aggregation activities spread out in different directions – which in turn increases the tendency of "gaps" to form in the aggregated knowledge. Any software solution for divergent phenomena should naturally support an element of *cognition*, that can fill gaps by generalizing on observed patterns. This is different from *deductive* capabilities that knowledge bases typically support, that derive implications based on inference rules.

To the best of his knowledge, the author is not aware of any approach to knowledge aggregation that addresses divergent phenomena.

3 Many Worlds on a Frame

To address the divergent problem of capturing operational knowledge, a data model called Many Worlds on a Frame (MWF) was developed. This data model is augmented with a set of cognitive heuristics that generalize on patterns of activities and suggest suitable gap-filling knowledge elements.

Although it was conceived and developed independently, the basic structure of the MWF model was found to be expressible as a superposition of two modal Frames in Kripke Semantics [5,6]. Comparison of MWF with Kripke Semantics can be found in [1]. The contribution of this work is hence not so much in the formal model itself, but in the way Kripke semantics can be used to address the divergent phenomena of operational knowledge aggregation.

Frame. The knowledge base is built on a global data structure called the *Frame*. The Frame represents a structure on which several interconnected "Worlds" can be described. The placement of different worlds on the Frame determines what is visible in each world, how changes in one world affect other worlds and what privileges do different users enjoy in each world. Formally, the Frame is defined as:

$$\mathsf{F} = (C, \rhd, \blacktriangleright, P, \sqsubseteq) \tag{1}$$

Here, C is a set of *Concepts* or *Worlds*, which are the building blocks of the knowledge base, and $\rhd$ and $\blacktriangleright$, subsets of $C \times C$, are reflexive, anti-symmetric and transitive relations representing conceptual generalization and containment respectively. P represents a set of user *privileges* and $\sqsubseteq$ represents a reflexive, anti-symmetric and transitive relation over $P \times P$ denoting a poset of user privileges.

Knowledge in MWF worlds are built from self-standing elements (known as *simpliciter* elements, in philosophy) called *concepts*. Elements of a given MWF world, including the world itself, is a concept. Concepts are organized in hierarchies, represented as rooted, acyclic graphs (partial orders with a greatest element) that form the Frame. Every concept belongs to two such hierarchies: the "is-a" or concept hierarchy and the "is-in" or containment hierarchy. A relation of the form $a \rhd b$ denotes that a "is a kind of" b and a relation of the form $a \blacktriangleright b$ denotes that a "is contained in" b.

The concept hierarchy inherits properties (and in the case of worlds – defined formally further down – the world structure). The containment hierarchy is used to inherit privileges and manage visibility. The root of the concept hierarchy is a concept called *Concept*, and the root of the containment hierarchy is a concept called $Universe\ of\ Discourse$ (UoD). *Concept* and UoD are related as follows:

$$\begin{aligned} UoD &\rhd Concept \\ Concept &\blacktriangleright UoD \\ UoD &\blacktriangleright UoD \\ Concept &\rhd Concept \end{aligned}$$

A concept that cannot be subclassed using the "is-a" relation, is called an *Instance* or a *Record*. Instances form leaf level concepts in the concept hierarchy, but not all leaf level concepts need be instances.

Each user in the system is assigned a *privilege* for any given concept. Privileges can be either explicitly assigned, or *derived* from other concepts. Privilege derivation happens through the "is-in" hierarchy. A user having a privilege $p \in P$ for a given concept $c \in C$ is guaranteed to have a privilege of at least p in all concepts that are contained in c.

World structure Each concept in a MWF system can also act as a local "World" and host a set of knowledge fragments in the form of associations across concepts. Formally:

$$\mathsf{w} = (C_w, \rho) \tag{2}$$

Here $\mathsf{w} \in C$, $C_w \subseteq C$ is a set of concepts that are said to be "imported" into w and ρ denotes a set of associations.

Any concept can be imported into any world, with some restrictions based on visibility – explained further down in this section. However, some concepts are imported by default into some worlds, and cannot be "un-imported" from them. The conditions for default imports are as follows:

1. A concept is considered to be imported in its own world:

$$\forall \mathsf{w}, \mathsf{w} \in C_w$$

2. All concepts contained in a world are imported by default into the world:

$$\forall x, x \blacktriangleright \mathsf{w} \Rightarrow x \in C_w$$

3. If a concept c is imported into another world w, all concepts contained in c are also imported into w:

$$c \in C_w \Rightarrow \forall x, x \blacktriangleright c, x \in C_w$$

The term ρ defines a set of associations among concepts in C_w. Associations are triples of the form $(source, predicate, target)$. Here *source* and *target* are concepts in C_w and *predicate* is a label describing the association.

In any association contained in world w, if the *target* concept is w itself, such associations are called *Roles*. The *source* concept is said to be playing a role defined by the *predicate* label in w.

Worlds can be created and added to the Frame in the following ways:

1. As an independently created *simpliciter* concept, with explicitly specified concept and containment parents
2. By *inheriting* from a world already existing in the Frame. This will inherit the world structure of the superclass (imported concepts and contained associations) into the subclass. Any predicate defined in the subclass will override the predicate of the same name (if any) inherited from the superclass
3. By *cloning* an existing world. This will create a new "sibling" world with the same "is-a" parent, while the "is-in" parent can be chosen by the user creating the clone. All imported concepts and associations among them will be replicated in the clone
4. By *inducing* a new world within another world w. This is done by choosing a set of concepts in C_w and invoking a world induce function. This will create a new world w' that is contained in w, whose "is-a" parent can be chosen by the user. The induce operation will also import the *induced subgraph* formed by all the chosen concepts from w and associations between them, into the new world w'.

If a world is a record (that is, a concept that is not subclassable), then users can also add *association instances* to the world in addition to associations. Association instances are again triples of the form $(source, predicate, target)$, where both *source* and *target* are records, rather than concepts. An association instance of the type (s, p, t) can be added to a world only if an association p connecting the types of s and t is already defined and visible in the world.

Addition and deletion of instances are called *data level* operations, while changes to or across concepts are called *structure level* operations. Changes to the "is-a" and "is-in" lineage of a world are called *Frame level* operations.

Privileges and Visibility. By definition, operational knowledge is subjective in nature. User privileges are an intrinsic part of the knowledge representation, rather than a feature of the application. User privileges can be broadly classified as follows:

Frame level privileges. These pertain to operations involving placement of a world in the Frame. This entails identifying and/or modifying its conceptual and containment parents

World structure level privileges. These pertain to operations that alter a world structure. This involves importing concepts into the world and adding roles and associations within the world

Data level privileges. These pertain to operations involving addition and deletion of association instances within a world – in other words, data elements that conform to a world structure

Membership grant and revoke privileges. These pertain to operations involving addition and deletion of members from worlds.

A world that is *protected* is invisible to non-members. A protected world c cannot be imported into another world w, if there is at least one member (explicit or derived) of w for whom c is not visible. Roles and associations in which the protected world c participates in w are not visible for non-members of w.

Cognitive gap-fillers. A divergent process of aggregation ends up forking off in different directions from a common knowledge pool and is susceptible to several knowledge *gaps.*

Any software environment to support divergent aggregation should contain an element of "*cognition*" that can generalize on observed patterns and suggest gap fillers. Various kinds of heuristics have been explored in the current implementation to act as gap fillers. Some example heuristics are explained below.

Data level heuristics. These are heuristics that observe patterns in which data elements participate in a world, and suggest gap fillers:

Principle of locality of relevance. This heuristic says that instances that participate in a world are typically found in the vicinity of the world itself. For instance, consider a world called *Department* contained in a larger world called $University$. Both *Department* and $University$ have imported the concept called $Person$. The principle of locality of relevance says that if a $Person$ instance is needed to play a role in the *Department* world, then the relevant $Person$ instance would typically already be featured in the $University$ world

Birds of a feather principle. This heuristic says that instances that play similar roles in a given world, tend to play similar roles in similar worlds. A common source of finding similar worlds are sibling worlds that share a common "is-a" ancestor in the Frame.

Structure level heuristics. These are heuristics that identify potential gaps in the structure of a world. These include suggestion of possible missing concept imports and/or associations:

Triadic closure. This heuristic says that if a concept a is associated with concepts b and c, the greater the "strength of the association" by virtue of the number of association instances, the greater the possibility that concepts b and c themselves are semantically related.

Clustering semantics. This heuristic says that concepts tend to operate in similar groups to describe semantics across different worlds. Suppose in a world w imported concepts $\{c_1, c_2, \ldots, c_n\}$ participate in several associations among themselves to form a tightly connected cluster. Now, if in another world w' a large part of the above set of concepts are imported, then it is likely that the rest of the concepts will also need to be imported in w' to build its structure.

All the above heuristics have an element of locality. But there are likely to be heuristics that can help connect seemingly disparate parts of the global knowledge. Work is still underway to explore the existence of such heuristic principles.

4 Conclusions

This paper introduced the divergent problem of aggregating operational knowledge in community settings, in contrast to the convergent problem of managing encyclopedic knowledge. A data model called MWF is also proposed to aggregate knowledge from such divergent phenomena.

The utility of operational knowledge is more immediate and concrete than the utility of encyclopedic knowledge. But it is unclear what would make users actively contribute knowledge to the knowledge base? What kind of *payoff mechanisms* can be associated with such utility? To answer these questions, we need to be looking into social science and group psychology. The MWF implementation RootSet is planned to be tested out in several kinds of community environments to study their dynamics and devise appropriate incentive mechanisms for users to actively contribute knowledge.

References

1. Srinivasa, S.: Many worlds on a frame: Aggregating operational knowledge in community settings. Technical report, Open Systems Laboratory, International Institute of Information Technology - Bangalore (2012)
2. Bon, G.: The Crowd. Transaction Publishers (1995)
3. Raafat, R.M., Chater, N., Frith, C.: Herding in humans. Trends in Cognitive Sciences 13(10), 420–428 (2009)
4. Razoumnikova, M.O.: Functional organization of different brain areas during convergent and divergent thinking: an eeg investigation. Cognitive Brain Research 10(1-2), 11–18 (2000)
5. Garson, J.: Modal logic. In: Zalta, E.N. (ed.) The Stanford Encyclopedia of Philosophy, Winter 2009 edn. (2009)
6. Woo, C.: Kripke semantics (version 17), `PlanetMath.org`, Freely available at `http://planetmath.org/KripkeSemantics.html`