The Hitch Hiker's Guide to the Galaxy

A.L. Cohen and S.B. Christiansen
IntelliMedia 2000+
Institute for Electronic Systems, Aalborg University
DK–9220 Aalborg Ø, Denmark
adam.cohen@bcs.org.uk, sbc@kom.auc.dk

> "In many of the more relaxed civilizations on the Outer Eastern Rim of the Galaxy, the Hitch Hiker's Guide has already supplanted the great Encyclopedia Galactica as the standard repository of all knowledge and wisdom, for though it has many omissions and contains much that is apocryphal, or at least wildly inaccurate, it scores over the older, more pedestrian work in two important respects. First, it is slightly cheaper; and secondly it has the words DON'T PANIC inscribed in large friendly letters on its cover."
>
> *Douglas Adams, The Hitch Hiker's Guide to the Galaxy (1979)*

1 Introduction

The *Hitch Hiker's Guide to the Galaxy* is a science fiction story written by Douglas Adams that foresees an electronic guide to Life, the Universe and Everything. Together with *The Digital Village*, Douglas Adams's company, we define the capabilities of a modern version of *The Hitchhiker's Guide*: providing information in a contextual setting; using a third party interaction metaphor via software agents; and providing personality-based feedback to encourage the user to trust the agents.

Since a complete *Hitchhiker's Guide* is not possible with current technology, we prototype a Guide that can provide guidance over a more limited domain. Our Guide answers natural language queries about places to eat and drink with relevant stories. These stories are created by *Storytelling Agents* from a knowledge base containing previously written reviews together with more detailed information about the places and the food and drink they serve.

Providing Information in Context

The Internet could be viewed as similar to the fictional *Hitch Hiker's Guide to the Galaxy*, at least in the apocryphal and wildly inaccurate senses. Unfortunately, it is missing the all-important, large friendly letters on the cover.

There are many Web sites that offer searching facilities. To use them, you type in some keywords and the search sites give you a list of sites that contain the words. The next stage is to flip to and from the list of links until you find the information you want. There are three problems occurring here:

1. There is no way to tell from the page of links which page contains the exact information you are looking for.
2. The information you are looking for might be spread across several pages.
3. The pages gathered by the search engine have no semantic links between them.

According to Roger Schank (Schank and Abelson 1977), the human mind operates in terms of context. We understand things only as they relate to experiences. Schank therefore proposed a model of human memory and understanding based on *scripts*, the stories we distill from our experiences. He suggested that we understand information by fitting it into scripts or by creating new ones. In order to do this, however, the information must be provided in context, not as lists of unrelated items.

Agent Technology

Most current day software interfaces use the *direct manipulation* metaphor. This implies both that the user can see what choices are available at any time and that the user is entirely responsible for making those choices. This metaphor breaks down when the user must handle the large amounts of information available on the Internet.

An alternative metaphor was proposed by the Knowledge Navigator (Apple Computer 1992). This vision of future computing allowed the user to *delegate* his tasks to the computer: instead of conducting a search by carefully selecting keywords, dates and author names, the user said, "Find me the paper I read on deforestation published by uh Flemson or something." The resulting loss of direct control necessitates that the user *trusts* the agent to complete its tasks satisfactorily.

An agent can gain the user's trust by gradually learning from the user's actions and reactions. In this way the user can see the agent developing and is "given time to gradually build up a model of how the agent makes decisions, which is one of the prerequisites for a trust relationship" (Maes 1994). Software agents learn the user's habits by building up a model of how they expect the *user* to behave. Two of the problems associated with such models are:

1. The agent must be able to convey the state of its model to the user so that she can see how the agent is developing.

2. A model can never fully capture the user's decision making process. For example, if a restaurant guidance agent is familiar with the restaurants that *you* like, what happens when you want one that's suitable for your parents?

Personality-based Expression

Communicating a complex internal model to the user is a difficult task. However, agents can harness people's expectations by using facial expressions to convey their internal state (how they "feel"). If done consistently, the user will quickly build a model of the agent's decision processes.

Naoko Tosa's *Neurobaby* (Tosa 1993) used neural networks to model an artificial baby that reacted in emotional ways to the sounds made by a user looking into its crib. *Neurobaby* was a fairly simple system, but because of the users' knowledge and expectations of a *real* baby, it was judged very convincing.

The concept of personalities can also counteract the second problem above. *Amalthaea* (Moukas 1996) is an information filtering and discovery system that makes use of evolving populations of genetically-based agents (each a separate "personality"). While the selection process ensures that the agents become customised to an individual user, the reproduction and mutation processes provide diversity in the population and in the resulting information.

Proposal

While there have been several thought experiments that combined the concepts in the previous three sections (for example, Apple's *Knowledge Navigator*), no one has yet come up with a satisfactory working product.

It is our belief that a successful presentation of information from varied sources must be both *relevant* and *easily understandable.* From our research we have discovered that these two criteria could be satisfied in a system that combines the three concepts of:

1. providing information in a contextual setting;

2. changing the interaction metaphor from direct to third party interaction;

3. and providing personality-based feedback to encourage the user to trust that third party.

Our primary aim in this project is to design and construct a Guide that successfully combines these three concepts.

2 Description of the The Guide

The Digital Village's *Principles of Guidance* (TDV and AT&T 1997) was our main inspiration for this project. This uses an in-depth profile of the user to answer natural language queries; it works by itself while the user is doing something else; and it can even use several methods of communication to keep in contact with the user.

Our Guide is not so advanced. While it does accept natural language queries and does answer in the form of stories, it does not work offline and only uses a simple (but still effective) profile of the user. It also has an interface that combines the modalities of speech input, text and graphics display, pointing input (mouse-based in our case, but pen-based on a mobile device) and positional input (such as Global Positioning System signals). We designed this interface to be scaleable from a desktop computer to a mobile phone.

Information Flow

Figure 1 summarises how the user's query is interpreted and processed by The Guide system. Briefly, the syntax and semantics of the user's query are analysed to determine her intention. This intention is then sent to the Storytelling Agents, which select and order the information from the FramerD database. This process results in *smarticles* which are displayed to the user. The sections that follow provide more details on each part of the system.

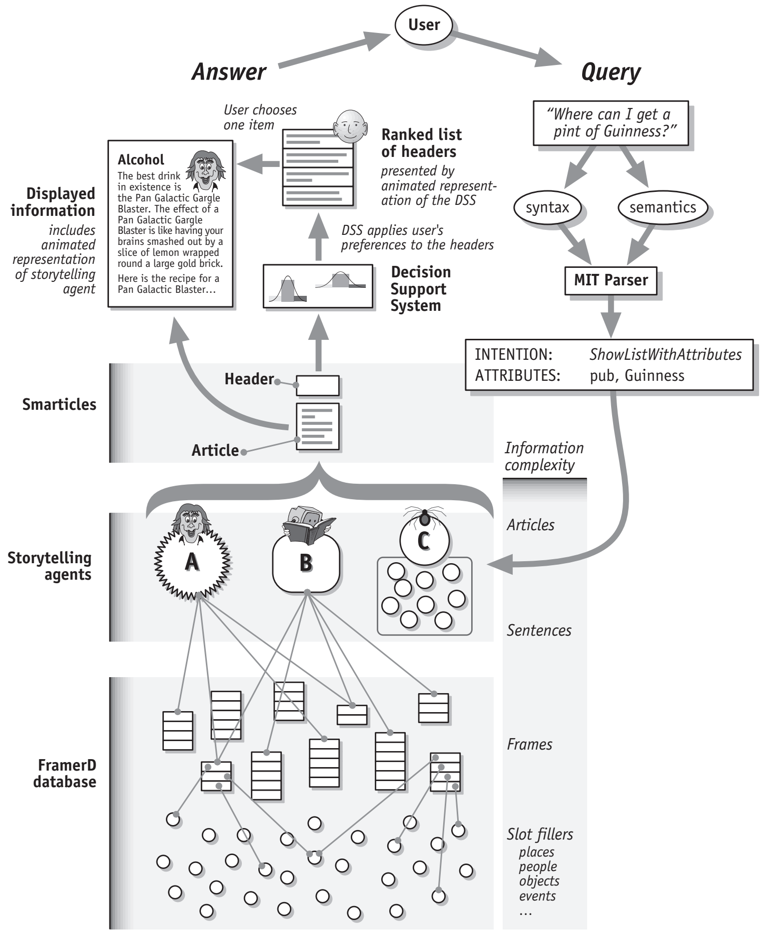

Figure 1: *Information flow in The Guide.*

FramerD Database

Since The Guide is a storytelling system, frames and scripts (Schank and Abelson 1977) provide a convenient way to represent its knowledge. FramerD (Haase 1996) is an object-oriented database designed to store and index frame-based information. It also comes with a large language knowledge base (KB) called *Brico*, consisting of the *WordNet* lexical thesaurus, the top-level *CYC* ontology (Cycorp 1999) and a semantic network of Roget's thesaurus. To this we added frames representing knowledge of the world (see Section 2 below).

Storytelling Agents

Having several storytelling agents provides The Guide with a greater diversity of information to present to the user. This counteracts the problem noted in Section 1 that the user model could never fully capture the user's decision process. All the agents can access the FramerD database to create their stories. Three possible types of storytelling agent are shown in Figure 1:

(A) Actual Human Being (AHB) agent: could be a human operator writing smarticles to answer the user's queries or a software agent that selects smarticles from a database of human-written ones.

(B) Predefined agent: has some fixed code that is executed to generate a smarticle each time it gets a query from the user.

(C) Genetically-based agent: an evolving population of agents similar to those in *Amalthaea* (Moukas 1996). Each of the agents in the population has a genotype that determines how it creates smarticles. The genotypes evolve according to a selection process based on whether the user chooses to read the smarticles. The genetically-based agent thus gradually adapts to the user.

The Guide currently has an AHB agent; an agent that uses the *Dada engine* (Bulhak 1996) to talk feasible nonsense; and an agent that interprets the user's query in terms of its own bias and returns combinations of existing smarticles.

Smarticles

In The Guide, smarticles are the stories that the storytelling agents tell. Each smarticle contains a header containing the attributes it involves so that it can be rated by the Decision Support System (DSS). The headers also contain a title and a reference to the authoring agent so that the available smarticles can be listed for the user to choose.

Decision Support System (DSS)

The available smarticles are sorted according to the penalty values they receive from the DSS. This uses a probabilistic user model that represents the user's likes and dislikes of different attributes. The Guide considers two kinds of attributes: ordered, where the attribute has a scalar value (e.g. distance); and unordered (e.g. type of food served). Each ordered attribute has an equivalent preference distribution in the DSS which represents how much the user dislikes that attribute. The unordered attributes are dealt with by a single preference distribution representing the user's least favourite unordered attribute.

The penalties are calculated using decision theory and the smarticles with the least penalties (those that best fit the current user model) are placed at the top of the list. The DSS updates the user model using Bayes's theorem each time the user chooses a smarticle to read. This probabilistic user model can quickly adapt to the user's preferences without the user having to provide any feedback other than choosing which smarticle she wants to read.

Graphical Display

The Guide has two main displays for presenting information:

- the list of smarticles, presented by an animated representation of the DSS (the *Presentation Agent*)
- and the smarticle the user chooses, presented by the authoring agent.

The Presentation Agent has emotions that reflect the state of the user model. The greater the accuracy of the user model (as measured by the success of the DSS's prediction of the user's choice), the less confused and the more happy it gets. This change in facial expression can be seen in Figure 2.

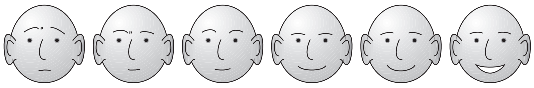

Figure 2: *The expressions of the Presentation Agent, from "Hopeless" to "Happy".*

Queries

TDV's *Principles of Guidance* (TDV and AT&T 1997) envisions complex natural language queries: if the user says he wants to sell Peril-Sensitive sunglasses, the Guide doesn't look for information about the glasses themselves but for people who would be interested in buying them. This involves the Guide understanding the user's *intention* rather than reducing the user's query to keywords.

However, intentions cannot be observed directly, but can only be inferred and recognised from the syntax, semantics and context of the communication. The system must have a deep knowledge of the world to be able to determine, for example, that Peril-Sensitive sunglasses might be worn by people interested in dangerous sports. It must also have a deep knowledge of language, since the same intention could be expressed in many different ways.

We use a parser developed by the Machine Understanding Group at MIT (Haase 1997) to help solve the second part of this problem. This parser uses the extensive language knowledge represented by *Brico* and is able to draw analogies between sentences.

Domain Knowledge

Providing knowledge of the subject (the *domain* knowledge) is a much harder problem. A Guide that could answer questions on Life, the Universe and Everything would need to be able to reason about every possible subject in great detail and such a knowledge base is not available. A Guide that knows a large amount about a particular subject is possible, however, and we therefore limit The Guide's domain to guidance on places to eat and drink.

To allow us to add new domains at a later stage, we designed a structure that could handle information on any subject, based on the Universal Context Model[1]. The Guide's knowledge structure is based on Roger Schank's types (*Events, Places, People* and *Things*) and extends them to five categories of *Events, Entities, Places/Objects, Concepts* and *Relationships.* It also defines particular attributes for each type, including the attributes of *Trust* (the respect given to the source of the information) and *Veracity* (fact, fiction or speculation).

We then define frames within this structure to represent our specific world knowledge. The resulting world knowledge structure can be seen in Figure 3.

Intention Analysis

Within a specific domain, people have specific intentions. Mc Kevitt (1991) has categorised several intention types relevant for the consultancy domain. However, The Guide's domain of guidance is slightly different — a consultant is assumed to be telling the truth with respect to achieving some specific task. To discover some of the intention types in our domain, we conducted an informal survey of three people. We asked them to think of all the questions that they might ask a system that knew everything there was to know about restaurants and pubs. The subjects were familiar with spoken dialogue systems but we asked them to imagine that they were talking to a person.

Our survey did not include a conversational context: the subjects were asked to think of individual questions and were not presented with any answers on which they could follow up.

[1] ©1998 Richard Harris, TDV.

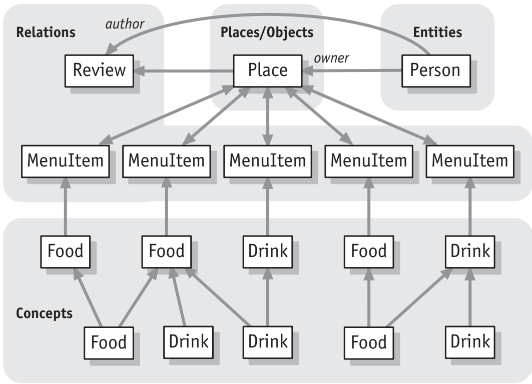

Figure 3: *The Guide's knowledge structure.*

Since our domain was different, we used Mc Kevitt's intention types as guidelines but not solutions. We then grouped the similar questions together and found there to be five main types of intention which we have labelled as follows:

Intention	**Example**
Show a list of places with specific attributes (ShowListWithAttributes)	"I want a free-range, organic meal in town."
Give details of a specific attribute at a specific place (DetailAttribute)	"What is the corkage price at *La Provence*?"
Give an opinion about a specific subject (GiveOpinion)	"What do you know about *Carlsberg Classic*?"
Show a list of places suitable for a specific situation (ShowListForSituation)	"I want to take my Grandma to a restaurant she'd like."
Generalisations (Generalisation)	"What percentage service do I have to pay in a Danish restaurant?"

Some of the intentions we found are similar to those in Mc Kevitt (1991) but there are significant differences due to the difference in domain. Notably, a guide does not provide a set of actions to achieve a goal, but instead provides opinions on possible actions that the user already knows about.

Because our survey did not include a conversational context, we did not find any discourse-related intentions such as repetition (repeated requests) or elaboration (requesting more information) (see Mc Kevitt 1991). This was a limitation of our survey and not of The Guide — it is designed to be able to notice and act upon sequences of intentions.

Restriction of User Input

Initially, we have restricted the intentions that The Guide detects to the simpler ones: ShowListWithAttributes, DetailAttribute and GiveOpinion. We then specified the information that each of these intentions should carry by designing frames such the one for a ShowListWithAttributes intention shown below:

ShowListWithAttributes	
INTENTION:	ShowListWithAttributes
UTTERANCE:	"I want a quiet, French meal and I don't mind driving."
ATTRIBUTES:	quiet, French, drive
USERLOC:	(9.9872632 E, 57.0138211 N, 126.2923482 m)
USERTIME:	3/5/1999 11:23:23.091 am

3 Results

As of now, we have only conducted module and integration tests of The Guide. By putting the system together we have achieved our goals as set out in Section 1.

Providing information in a contextual setting: The Guide tells stories (albeit simple ones) in response to the user's queries rather than giving a list of unrelated information.

Third party interaction: the user interacts with the system via the presentation agent, delegating the search to the system rather than specifying keywords herself.

Providing personality-based feedback: the state of the system's probabilistic model of the user is conveyed back to the user by an animated humanoid face.

The following paragraphs decribe the operation of The Guide as seen in Figure 4:

Screen 1: When The Guide is started, the user is presented with a welcome screen. She can then either speak or type her question to the guide.

Screen 2: The **Presentation Agent** then indicates that the system is busy processing. Meanwhile, The Guide attempts to interpret the user's intention using a combination of the MIT parser and word-spotting. If the intention is recognised, The Guide builds an intention frame as in Section 2 and passes it to the **Storytelling Agents**.

Screen 3: Each **Storytelling Agent** is free to operate in its own way, using all the information in the **FramerD** database. Once they have generated their smarticles, they pass them back to the user via the **DSS**. The **DSS** sorts the smarticles according to its probabilistic user model and indicates its current state using the **Presentation Agent**.

Screen 4: The user can choose either to read one of the smarticles or to cancel her question. Choosing one of the smarticles results in a graphical representation of the authoring **Storytelling Agent** presenting the chosen smarticle. The **DSS** then updates its user model to reflect the new information provided by the user's question and her choice of smarticle. At this point, the user can cancel and go back to the list of smarticles, ask another question straight away or use the pointer to indicate that she wants to ask another question.

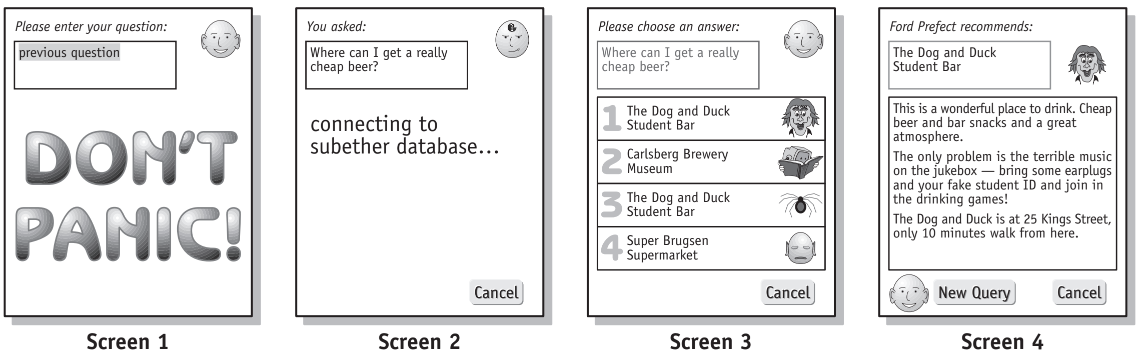

Figure 4: *The Guide: what the user sees.*

4 Conclusion

Compared to many of the large research projects in the field of Intelligent MultiMedia, The Guide is a simple prototype. However, as Schank (1990) points out, "a computer that had thousands or hundreds of stories and carefully selected which ones to tell might well be considered to be not only intelligent but wise" (p. 15). By using intention analysis and a probabilistic user model, The Guide can easily select relevant stories and only seems to lack a large collection of stories to be considered wise.

Possible future developments could include improving the storytelling capabilities of The Guide, perhaps by using a tailored version of the Dada engine to write believable reviews, or by using artificial life -based **StoryAgents** to provide a greater level of personalisation. The interface could be further personalised to the users by analysing the *sequence* of their intentions as well the intentions themselves.

Though there are many successful projects that provide guidance over limited domains, a Guide to Life, the Universe and Everything is still in the realm of science fiction. Fortunately, The Guide has the words DON'T PANIC inscribed in large friendly letters on its cover.

Acknowledgements

Richard Harris, Technical Director of *The Digital Village*, for providing us with the Universal Context Model and Principles of Guidance documents.
Prof. Ken Haase of the Machine Understanding Group at the MIT Media Lab for his help in compiling FramerD and for allowing us to use his unpublished parser software.
We would also like to thank our supervisor, Dr. Paul Mc Kevitt, for his invaluable advice, and Ruth Kushner (Adam's fiancée), for her support.

References

Adams, D. (1979). *The Hitch Hiker's Guide to the Galaxy.* London: Barker.

Apple Computer (1992, May). Knowledge navigator. In B. A. Myers (Ed.), *Conference on Human Factors in Computing Systems: Special Video Program*, Monterey, CA. ACM/SIGCHI.

Bulhak, A. C. (1996, April). On the simulation of postmodernism and mental debility using recursive transition networks. Technical Report 96/264, Dept Computer Science, Monash Univ., Melbourne, Australia.

Cycorp (1999, Apr). Cyc ontology. http://www.cyc.com.

Haase, K. (1996). FramerD: Representing knowledge in the large. *IBM System Journal 35*, 381–397.

Haase, K. (1997, October). Chopper, an unprincipled parser. Kindly provided by Prof. Haase.

Maes, P. (1994, July). Agents that reduce work and information overload. *Communications of the ACM 37*(7), 31–40.

Mc Kevitt, P. (1991). *Analysing coherence of intention in natural language dialogue.* Ph. D. thesis, Department of Computer Science, University of Exeter, England.

Moukas, A. (1996, April). Amalthaea: Information discovery and filtering using a multiagent evolving ecosystem. In *Proceedings of the Conference on Practical Application of Intelligent Agents and Multi-Agent Technology*, London.

Schank, R. (1990). *Tell Me A Story: a new look at real and artificial memory.* Charles Scribner's Sons.

Schank, R. and R. Abelson (1977). *Scripts, Plans, Goals and Understanding.* Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

TDV and AT&T (1997, August). Principles of guidance. Commercial-in-confidence.

Tosa, N. (1993). Neurobaby. In *SIGGRAPH-93 Visual Proceedings. Tomorrow's Realities*, pp. 212–213. ACM SIGGRAPH.