

Observing the User Experience

A Practitioner's Guide to User Research

Mike Kuniavsky



MORGAN KAUFMANN PUBLISHERS

AN IMPRINT OF ELSEVIER

SAN FRANCISCO SAN DIEGO NEW YORK BOSTON
LONDON SYDNEY TOKYO

CHAPTER 8

Contextual Inquiry, Task Analysis, Card Sorting

Although a Web site is many things—entertainment, advertising, information—every Web site is also a tool. It's a tool to help people do something that, otherwise, would be difficult, time consuming, expensive, or inefficient. Even if the efficiency it introduces is merely a shortcut to something funny or interesting, it's still serving as a kind of tool. It's something that makes their lives a little easier. It solves some problem.

Tools solve problems, and to build the right tool, you need to know what the problem is. There are a number of ways to do that. You can guess, using your knowledge of the target audience and what they're trying to do. This is fast, but it's fraught with danger: if you're not a member of the target audience (which, as a developer, you rarely are), your understanding of the nature and severity of your users' problems will not be the same as theirs. You could decide that someone needs a bigger hammer, when in fact, he or she needs smaller nails.

Another method is to ask representatives of the target audience what their problems are, but this too can lead you in the wrong direction. People tend to idealize their needs and desires, and their statements often don't correspond to their actual needs and behavior. Many sites have been perplexed about why no one uses their personalization functionality after everyone in a survey said they wanted it. The reason is simple: ideally, people would love to have everything tuned perfectly to their preferences and needs, but the actual tuning process is much harder than they first imagine. So although they

would love a service that's personalized as they imagine it, when faced with the task of having to make it that way themselves, they'd rather use the plain vanilla service.

Moreover, the obvious problem isn't always the real problem. The person who wants a new hammer and smaller nails? Maybe she really just needs a cheap birdhouse, so she's making one. Once she's done, maybe she'll never need the hammer and nails again. The best way to find out what people's real problems and needs are, and how they really do things, is to discover them for yourself. The techniques in this chapter are designed to reveal how your target audience lives, how they think, and what problems they run into.

These procedures are best done before specific solutions have been created, which is generally at the beginning of a development process. Ideally, they're done even before any technologies have been developed or ideas for solutions have been suggested. Then, all the development effort can go into solving the target audience's problems rather than trying to adjust existing technology to fit their needs. In practice, that rarely happens, but it's still good to do this research as early as possible.

Contextual Inquiry

The definitive book on *contextual inquiry*—and one that this chapter is deeply indebted to—is Hugh Beyer and Karen Holtzblatt's *Contextual Design: Defining Customer-Centered Systems*. They define this procedure as follows:

Contextual inquiry is a field data-gathering technique that studies a few carefully selected individuals in depth to arrive at a fuller understanding of the work practice across all customers. Through inquiry and interpretation, it reveals commonalities across a system's customer base.

In other words, contextual inquiry is a technique that helps you understand the *real* environment people live in and work in, and it uncovers their needs within that environment. It uncovers what people really do and how they define what is actually valuable to them. It can reveal unexpected competition (is it just other Web sites, or

are you competing with some real-world phenomenon?) and people's real values.

As a technique based in anthropology and ethnography, the basic method of research involves visiting people and observing them as they go about their work. In watching them carefully and studying the tools they use, it's possible to understand what problems people face and how your product can fit into their lives.

When Contextual Inquiry Is Appropriate

Ideally, as stated previously, every development cycle would start with a contextual inquiry process—not technologies, or solutions, or problem statements, or development teams. The researchers would pick a target audience and research them to uncover their needs. Maybe the biggest cause of their problems lies in the exchange and transportation of their data, but maybe it's not that at all. Maybe the users' biggest problem is that people spill a lot of coffee on their printouts. In an abstract development cycle, either problem would be acceptable to uncover and tackle. The first would lead to some kind of software solution. The second, to cupholders and laminated paper.

Since the focus of this book is information related, we will ignore the cupholder solution because there *are* practical limits to the problems that software designers can tackle, but it's useful to remember that the problems people encounter and their ultimate goals will not just be informational. The consideration of users' experiences should not be limited to their actions with the product. Their goals will have something to do with their lives, and that could be anything. Thus, from the users' perspective, a Web site that's about skateboarding tricks is not about sharing or sorting skateboarding trick information—that's a secondary purpose; it's about getting out on a skateboard and doing new tricks. The information and the site are tools that make finding trick instructions more efficient, but the real goal lies with the skateboard, not the Web site.

Most projects begin with an idea about what problems must be solved and also a rough idea about how to solve them. Contextual inquiry clarifies and focuses these ideas by discovering the exact situations in which these problems occur, what these problems entail,

and how people solve them. Thus, it's best done before the process of creating solutions has begun, which is most often the very beginning of the development cycle.

It can also be done in between development cycles or as part of a redesign. In those situations, it can tell you how people are using the product, when they're using it, and what they're using it for. This serves as a check of your initial assumptions and as a method of discovering areas into which the product can naturally expand (Table 8.1).

Timing	Activity
$t - 2$ weeks	Organize and schedule participants.
t	Begin interviews. Begin analysis scheduling process for development team.
$t + 1$ week	Complete interviews. Review videotapes and take notes. Complete analysis scheduling.
$t + 2$ weeks	Prepare space for group analysis. Verify participant schedule.
$t + 2$ weeks	Analyze affinity diagram (one day).
$t + 2$ weeks + 1 day	Begin results analysis. Write report and present to stakeholders.

The Contextual Inquiry Process

Since you'll be going out of your office and into the workplaces and homes of your customers, it's especially important to be thoroughly prepared. You won't have the option to go back and get something you've forgotten, and you'll be the one making the first impression about your company (not the cool new leather waiting room furniture).

Target Audience

Choosing an appropriate target audience is described in detail in Chapter 6 of this book, but the short version is that you should pick people like the people you think will want to use your product. Maybe they use the product already. Maybe they use a competitor's product. Regardless, they should have the same profile as the target audience that will eventually use what you have to offer.

You should specify this target audience in as much detail as you can, concentrating on their behavior.

- What is their demographic makeup?
- What is their Web use profile?
- What tasks do they regularly do?
- What tools do they regularly use?
- Are there tools they must occasionally use to solve specific problems?
- How do they use them?

Concentrate on the most important customers. Your product may appeal to a varied group of people, but there are only going to be a couple of key target markets. In fact, there may be only one. Focus all of your research there until you feel that you know what there is to know about their behavior, and then move on to secondary markets.

Recruiting

Once you have your profile, you need to find people that match it. A complete description of recruiting is in Chapter 6, but here are some things to consider when recruiting for contextual inquiry.

First, decide how many people you want to visit. The number will depend on how much time you have allocated to the research and the resources available. Beyer and Holtzblatt suggest observing 15 to 20 people, but that can be prohibitive because of the amount of interview and analysis time it requires. Five to eight people should give you a pretty good idea of how a big chunk of the target audience does their work (or entertains themselves, or shops, or whatever the focus of your product happens to be) and should be enough for a first round of inquiry. If you find that you have not

met the goals of the research or you don't feel comfortable with the results at the end of the first round, schedule a second round.

Scheduling

Once you've found your group of candidates, you need to schedule them. Contextual inquiry sessions can last from a couple of hours to a full workday, depending on the length of the tasks and how much ancillary information you'll be collecting. The most important criterion in scheduling is that the people need to be doing the kinds of activity you're going to study while you're observing them. It may be necessary to negotiate with them so you show up when they're planning to do the relevant tasks or to ask them to wait until you arrive.

Interviews can last from an hour to three or four. Schedule one or two interviews per day, along with some time to review your observations. If you have multiple observers, you can schedule several interviews at once (this makes some logistics easier, but requires extra equipment).

Since the research is going to be onsite, give the participants some idea of what to expect. Tell them the general goals of the research, how long it will take, what equipment is going to be used, and what kinds of things you're going to be looking at. You don't have to be specific (and, in fact, leaving some specifics out can produce a more spontaneous response), but they should have a good idea of what's involved before you show up. Tell them not to prepare for your arrival at all, that it's important that you see how they work, warts and all.

When studying people in office environments, it's often necessary to get releases and to sign nondisclosure agreements. Sometimes it's possible to do stealth research under the promise of anonymity, but video equipment is pretty visible, and it's hard to ignore a stranger hanging out all morning watching a certain cube and taking notes. If there's any doubt, ask the people you've scheduled to tell everyone who needs to know about your arrival and to get you all the forms you need to have as early as possible.

The incentive payment should reflect the length of the observation and should be between \$100 and \$200 for most research.

However, companies may have policies restricting such payments to their employees, which should be determined ahead of time (this is especially true when the company being studied is the same company commissioning the research—as is often the case for intranet or in-house software projects). If cash payment is not allowed, then a small gift may be appropriate (though not for government agencies or regulated industries). Follow-up interviews should be treated likewise unless you agree with the participants ahead of time on a single lump sum (in which case, it should reflect the total amount of time spent).

Learn the Domain

In order to be able to understand what people are doing and to properly analyze your data, you need to be familiar with what they do. This means familiarizing yourself with the terminology, the tools, and the techniques that they are likely to be using in their work. You don't have to know all the details of their job, but you should be somewhat familiar with the domain.

If you know nothing about a task, before you begin your research, you can have someone familiar with it walk you through a session. If you have the time, you can also use the “sportscaster method,” having one expert explain what another one is doing, as in a play-by-play commentary. They don't have to go into complicated technical explanations, just enough to familiarize you with the job.

If possible, try the task yourself. If it's something that is done with software, ask to borrow a copy (or a training manual) and use it for a couple of hours. If it's something that's done physically and you can try it without risk, ask to try it (this works for things like making pizza and data entry, but it's not as successful for things like nuclear chemistry). If the environment you're studying is a technical one, ask a member of technical support or quality assurance to walk you through some typical tasks to see how they, as expert in-house users, do them.

In general, the more you know about the tasks your target audience does, the better you'll be able to interpret their behavior when you observe it.

Develop Scenarios

Warning Be careful not to let your scenarios bias your observations. Use them as a framework to structure your interview, not to create expectations of how people do or don't behave in general.

As part of your preparation, you should be explicit about your expectations. Attempt to write down how and when you expect the people you're observing to do certain things that are important to your product, and what attitudes you expect they will have toward certain elements. You can do this with other members of the development team, asking them to profile the specific actions you expect people to take in the same way that you profiled your audience's general behavior in Chapter 7. This will give you a platform against which you can compare their observed behavior.

When you're in the interview, keep these scenarios in mind while watching people. Keep an open mind about what you're seeing, and use the situations where observed behavior doesn't match your expectations as triggers for deeper investigation.

Practical Stuff

In addition to all the contextual inquiry-related preparation, there are a number of things you should do just because you're leaving the comfort of your office:

- Make a list of everything you're going to bring—every pencil, videotape, and notebook. Start the list a week before you're going to do your first interview. Then, whenever you realize there's something else you should bring, add it to the list. A day before the interview, make sure you have everything on the list (I put a check next to every item I have) and get everything you don't. On interview day, cross off everything as it's loaded into your backpack or car.
- If you're going to get releases (either to allow you to observe or for the participants to participate), make sure you have twice as many as you expect to need.
- Bring everything you need for the incentive payment. This could be a disbursement form from accounting, a check, or an envelope with cash. Blank receipt forms are useful, too (though some accounting departments will accept a participant release as proof of receipt of payment).
- Know how to operate your equipment. Set up a test site that will simulate the user's work environment ahead of time. A day or two before, set up everything as you're going to use it

Note Sometimes real-life situations unfold very differently from how you may have expected. You may be expecting to find one kind of situation—say, a typical day using the typical tools—and you find something completely different, a crisis where the main system is down or they're scrambling to meet an unexpected deadline. In such situations, pay attention to how the unexpected situation is resolved and compare that to the situation you had expected and that which others experience. If the situation is totally atypical—it happens only every five years or the people you're interviewing have been pulled into a job that doesn't relate to a task you're studying—try to get them to describe their normal routine, maybe in contrast to what they're doing at the moment. If the situation seems like it's too far off from what you're trying to accomplish, reschedule the interview for a time where their experience may be more relevant to your research goals.

onsite, complete with all cords plugged in, all tripods extended, all cameras running, and all laptops booted. Then break it down and set it back up again. Get a set of good headphones to check the quality of the audio. Good audio quality is the most important part of the video-recording process.

- Have more than enough supplies. Bring an extension cord, two extra videotapes, two extra pads of paper, and a couple of extra pens. You never know when a tape will jam or an interview ends up being so exciting that you go through two notepads. If you plan on using a laptop to take notes, make sure that you bring either extra batteries or an AC adapter (one will suffice).
- Make plans for meal breaks. Closely watching someone for several hours can be draining, and you don't want to run around an office frantically looking for a drinking fountain while worried that you're missing a key moment. Bring bottled water and plan to eat in between sessions, or have lunch with your participants, which can be a good opportunity to get background on their jobs in a less formal setting.

Conducting the Inquiry

One of the keys to getting good feedback in a contextual inquiry situation is establishing rapport with the participant. Since you want to watch him or her working as "naturally" as possible, it's important to set out expectations about each other's roles. Beyer and Holtzblatt define several kinds of relationships you can strive for.

- The *master/apprentice* model introduces you as the apprentice and the person who you'll be watching as the master. You learn his or her craft by watching. Occasionally, the apprentice can ask a question or the master can explain a key point, but the master's primary role is to do his or her job, narrating what he or she is doing while doing it (without having to think about it or explain why). This keeps the "master craftsman" focused on details, avoiding the generalizations that may gloss over key details that are crucial to a successful design.
- *Partnership* is an extension of the master/apprentice model where the interviewer partners with the participant in trying

to extract the details of his or her work. The participant is made aware of the elements of his or her work that are normally invisible, and the partner discusses the fundamental assumptions behind the work, trying to bring problems and ways of working to the surface. The participant is occasionally invited to step back and comment about a certain action or statement and think about the reasons for his or her behavior. Although this can potentially alter the participant's behavior, it can also provide critical information at key points.

Beyer and Holtzblatt also point out several relationships to avoid.

- The *interviewer/interviewee*. Normally, an interviewee is prompted by an interviewer's questions into revealing information. Interviewees won't reveal details unless specifically asked. That's not the situation you want in a contextual inquiry situation. You want the participant's work and thoughts to drive the interview. When you find yourself acting as a journalist, prompting the participant before he or she says something, refocus the interview on the work.
- The *expert/novice*. Although you may be the expert in creating software for helping them, the participants are experts in their own domain. As Beyer and Holtzblatt suggest, "Set the customer's expectation correctly at the beginning by explaining that you are there to hear about and see their work because only they know their work practice. You aren't there to help them with problems or answer questions." They suggest that it should be clear that the goal is not to solve the problems then and there, but to know what the problems are and how they solve them on their own. If you are asked to behave as an expert, use nondirected interviewing techniques and turn the question around, "How would you expect it to print?"
- Don't be a *guest*. Your comfort should not be the focus of attention. You are there to understand how they do their work, not to bask in the light of their hospitality. Be sensitive to the protocol of the situation. If good manners dictate acting as a guest for the first few minutes of your visit, then do so to make the participants comfortable, but quickly encourage them to get on with their work and move into a more partnership-based dialogue.

Warning Sometimes management *may* want you to report on specific employees and their performance without telling them ahead of time that you'll be doing so. In such situations, explain the ethical problems with doing so—that it violates the confidentiality that your interview subject has placed with you—and that it may violate labor laws.

Note I'm using the word *action* to refer to a single operation during a task. In most cases, it's something that takes a couple of seconds and can be described with a single, simple idea. Actions are then grouped into *tasks*, which are things that satisfy a high-level goal. Task granularity can range all over the board. A task can involve something as straightforward as filling out a form or something as complex as picking out a car.

- Another role to avoid is *big brother*. You are not there to evaluate or critique the performance of the people you are observing, and that should be clear. If they feel that way, then they're not likely to behave in typical ways. Moreover, if participation in your research is at the request of their managers, it can seem that this is a sneaky way to check up on them. Emphasize that you are not in a position to evaluate their performance. If possible, once you've gotten permission from management to do so, contact and schedule people yourself rather than having it come as a demand from above.

Inquiry Structure

The structure of the inquiry is similar to the structure of most interviews, except that the majority of it is driven by the interviewee's work rather than the interviewer's questions. The structure follows the general interview structure described in Chapter 6: introduction, warm-up, general issues, deep focus, retrospective, wrap-up.

The *introduction* and *warm-up* should be times for the participant and the interviewer to get comfortable with each other and to set up expectations for how the observation will proceed. This is the time to get all the nondisclosure forms signed, describe the project in broad detail, and set up the equipment. Check that the image and sound recording is good, and then don't fuss with the equipment again until the interview is over since it'll just distract people from their work. Describe the master/apprentice model and emphasize your role as an observer and learner. Remind the participant to narrate what he or she is doing and not to go for deep explanations.

Once you're in position, you may want to ask some *general questions* to gain an understanding of who the person is, what his or her job is, and what tasks he or she is going to be doing. Ask the participant to provide a description of a typical day: what kinds of things does he or she do regularly? What are occasional tasks? Where does today's task fit into a typical day? Don't delve too deeply into the reasons for what he or she does; concentrate on actions and the sequence.

This will begin the *main observation period*. This phase should comprise at least two-thirds of the interview. Most of the time should be spent observing what the participants are doing, what

Warning Maintaining authenticity is a crucial part of observation. If you sense that the person you're watching is not doing a task in the way that they would do it if you were not watching, ask him or her about it. Ask whether he or she is doing it is how it should be done, or how it is done. If the former, ask him or her to show you the latter even if it's "really messy."

Note Provide privacy when people need it. Tell the people you're observing to let you know if a phone call or meeting is private—or if information being discussed is secret—and that you'll stop observing until they tell you it's OK to start again. Pick a place to go in such a situation (maybe a nearby conference room or the cafeteria) and have them come and get you when they're finished.

tools they are using, and how they are using them. Begin by asking them to give a running description of what they're doing, as to an apprentice—just enough to tell the apprentice what's going on, but not enough to interrupt the flow of the work—and then tell them to start working. As an apprentice, you may occasionally ask for explanations, clarifications, or walk-throughs of actions, but don't let it drive the discussion. I find that taking occasional notes while spending most of my energy concentrating on their words and actions works well, but it requires me to watch the videotape to get juicy quotations and capture the subtlety of the interaction. Others recommend taking lots of notes onsite and using the videotape as backup. Regardless, you should have a clear method to highlight follow-up questions. I write them in a separate place from the rest of my notes.

When either the task is done or time is up, the main interview period is over. An immediate *follow-up interview* with in-depth questions can clarify a lot. Certain situations may not have been appropriate to interrupt (if you're observing a surgeon or a stock trader, that may apply to the whole observation period), whereas others may have brought up questions that would have interrupted the task flow. As much as possible, ask these while the participant's memory is still fresh. To jog people's memories, you can even rewind the videotape and show them the sequence you'd like them to describe in the viewfinder (but only do this if you can find it quickly since your time is better spent asking questions than playing with the camera). To quote Victoria Bellotti, senior scientist at Xerox PARC, "You'll never understand what's really going on until you've talked to people about what they are doing. The [follow-up] interview . . . gives you the rationale to make sense of things that might otherwise seem odd or insignificant." If there are too many questions for the time allotted, or if they're too involved, schedule another meeting to clarify them (and schedule it quickly, generally within one or two days of the initial interview since people's memories fade quickly).

Wrap-up the interview by asking the participant about the contextual inquiry experience from his or her perspective. Was there anything about it that made him or her anxious? Is there anything the participant would like to do differently? Are there things that you, as the apprentice, could do differently?

Beyer and Holtzblatt summarize the spirit of the interviewing process as follows:

Running a good interview is less about following specific rules than it is about being a certain kind of person for the duration of the interview. The apprentice model is a good starting point for how to behave. The four principles of Contextual Inquiry modify the behavior to better get design data: *context*, go where the work is and watch it happen; *partnership*, talk about the work while it happens; *interpretation*, find the meaning behind the customer's words and actions; and *focus*, challenge your entering assumptions.

So what you want most is to come in unbiased and, with open eyes and ears, learn as much as you can about how the work is done while trying to find out why it is so.

What to Collect

There are four kinds of information you should pay attention to when observing people at work. Each of these elements can be improvised or formal, shared or used alone, specific or flexible.

- The *tools* they use. This can be a formal tool, such as a specialized piece of software, or it can be an informal tool, such as a scribbled note. Note whether the tools are being used as they're designed, or if they're being repurposed. How do the tools interact? What are the brands? Are the Post-its on the bezel of the monitor or on the outside flap of the Palm Pilot?
- The *sequences* in which actions occur. The order of actions is important in terms of understanding how the participant is thinking about the task. Is there a set order that's dictated by the tools or by office culture? When does the order matter? Are there things that are done in parallel? Is it done continuously, or simultaneously with another task? How do interruptions affect the sequence?
- Their *methods* of organization. People cluster some information for convenience and some out of necessity. The clustering may be shared between people, or it may be unique to the individual being observed. How does the target audience organize the

- information elements they use? By importance? If so, how is importance defined? By convenience? Is the order flexible?
- What *kinds of interactions* they have. What are the important parties in the transfer of knowledge? Are they people? Are they processes? What kinds of information are shared (what are the inputs and outputs)? What is the nature of the interaction (informational, technical, social, etc.)?

The influences of all four of these things will, of course, be intertwined, and sometimes it may be hard to unravel the threads. The participant may be choosing a sequence for working on data, or the organization of the data may force a certain sequence. Note the situations where behaviors may involve many constraints. These are the situations you can clarify with a carefully placed question or during the follow-up interview.

Artifacts

Artifacts are—for the purposes of contextual inquiry—the nondigital tools people use to help them accomplish the tasks they're trying to do. Documenting and collecting people's artifacts can be extremely enlightening. For example, if you're interested in how people schedule, it may be appropriate to photograph their calendars to see what kinds of annotations they make, or to videotape them using the office whiteboard that serves as the group calendar. If you're interested in how they shop for food, you may want to collect their shopping lists and videotape them at the supermarket picking out items. It's doubtful that you'd want to collect a surgeon's instruments after an operation, but you may want a record of how they're arranged. Having a digital camera with a large storage capacity can really help artifact collection and digital pictures of artifacts, and how people use them can make great illustrations in reports and presentations.

Always make sure to get permission when you copy or collect artifacts.

Here is a snippet of the notes from an observation of a health insurance broker using an existing online system to create an RFP (request for proposal):

Looks at paper [needs summary] form for coverage desired. Circles coverage section with pen.

Goes to Plan Search screen.

Opens "New Search" window.

"I know I want a 90/70 with a 5/10 drug, but I'm going to get all of the 90/70 plans no matter what."

Types in plan details without looking back at form.

Looks at search results page.

Points at top plan: "Aetna has a 90/70 that covers chiro, so I'm looking at their plan as a benchmark, which is enough to give me an idea of what to expect from the RFP."

Clicks on Aetna plan for full details.

Prints out plan details on printer in hall (about three cubes away) using browser Print button. Retrieves printout and places it on top of needs summary form.

Would like to get details on similar recent quotes.

Goes back to search results. Scrolls through results and clicks on Blue Shield plan.

Video Recording

I recommend videotaping every contextual inquiry (in fact, every interview you have, period) since it provides an inexpensive record that allows for more nuanced analysis than any other method. Some key events happen only once. Notes take the interviewer/moderator/observer's attention away from the participants and may not be able to capture all the details that make the event important. Two-person teams can be intimidating, and require two skilled people to be present for all interviews. Audio recording doesn't record body language or facial expressions, which can add vital information to what people really think and feel. Only video can approach the full experience of being there. At worst, it's no worse than audio; at best, it's much better.

However, recording good interviews on video takes some practice, especially when working in the field. The twin dragons that can keep you from a good recording are *insufficient light* and *noisy sound*.

(Continued)

- **Light.** Most video cameras aren't as sensitive as human eyes, so a room that looks perfectly fine to the human eye can be too dark for a run-of-the-mill video camera. When there's not enough light, you don't capture the details of the environment. Fortunately, recent generations of consumer cameras have enough low-light sensitivity to be useful in all but the dimmest environments. If you anticipate having to tape in a dark environment, try to find a camera that can work in 2 lux or less. This will give you enough sensitivity that even things in partial shadow can be videotaped. Be aware of different light levels and the way your camera compensates for them. A bright computer monitor in a dim room can cause the camera to make the rest of the image too dark. A picture window onto a sunny day can cause everything inside a room to appear in shadow. Many cameras can compensate for these situations but need to be set appropriately.
- **Sound.** The hum of a computer fan. The hiss of an air conditioner. Laughing from the cafeteria next door. It's amazing what ends up recorded instead of what you wanted. Good sound is probably the most technically important part of the whole recording process. Getting it can be tricky. Most video cameras have omnidirectional microphones that pick up a broad arc of sound in front (and sometimes behind) the camera. This is great for taping the family at Disney World, but not so great when you're trying to isolate a participant's voice from the background noise. Get a camera with a good built-in microphone, an external microphone jack, and a headphone jack. Get a decent set of headphones, plug them into the headphone jack, and walk around with the camera, listening to how it picks up sound. Then, when you're onsite, adjust the camera with the headphones to minimize how much external noise it picks up. You may have to do this at the cost of getting a good image, or you may have to get an external directional microphone and mount it separately. If worse comes to worst, you can use an external lapel microphone on a long cord. Keep the headphones attached to the camera throughout the interview and discreetly check the audio quality every once in a while.

Even the best equipment is sometimes insufficient or superfluous. Some situations are impossible or awkward to videotape. In others, you can't get permission. In those cases, audio recordings can often be used, but they carry only part of the information and tend to be more labor intensive to analyze. Use audio recording when you can, but be prepared to fall back on paper and pen.

How to Analyze Contextual Inquiry Data

The output from Contextual Inquiry is not a neat hierarchy; rather, it is narratives of successes and breakdowns, examples of use that entail context, and messy use artifacts.

—Dave Hendry, Assistant Professor, University of Washington
Information School, personal email

How data should be interpreted should differ based on the task under examination. Contextual inquiry helps you understand how people solve problems, how they create meaning, and what their unfulfilled needs are. It does this largely through six methods.

- Understanding the *mental models* people build. People don't like black boxes. They don't like the unknown. They want to understand how something works in order to be able to predict what it will do. When the operation of a process or tool isn't apparent, people create their own model for it. This model helps them explain the results they see and frames their expectations of how it will behave in the future. This model may have nothing to do with the actual functionality of the tool, but if it matches their experience, then it's the one they're going to use. Knowing the model being used allows you to capitalize on that understanding and meet people's expectations.
- Understanding the *tools* they use. Since you're building a tool that's supposed to replace the tools people use now, it's important to know what those tools are and how they're used. Rather than leisurely browsing a catalog as you would expect them to do, they may just check the specials page and let an online comparison engine find the cheapest price. They may keep addresses in a Palm Pilot, a carefully organized address book, or they may just have a pocketful of scribbled-on scraps of paper. Maybe your competition isn't Outlook or the Day-timer, but napkins!
- Understanding the *terminology* they use to describe what they do. Words reveal a lot about people's models and thought processes. When shopping for food, people may talk in terms of meals, calories, or food groups. They may use one ("bread") to represent another ("carbohydrates") or misuse technical terminology (using "drivetrain" to talk about a car's suspension, for example). Paying attention to the words people use and the way they use them can reveal a lot about their thought patterns.
- Understanding their *methods*. The flow of work is important to understanding what people's needs are and where existing tools are failing them. Unraveling the approach people take to solving a task reveals a lot about the strengths and weaknesses

- of the tools they use. If someone composes a message in a word processor and then cuts and pastes it into an email program, that says something about how he or she perceives the strengths of each product. If he or she is scribbling URLs on a pad while looking through a search engine, it says something about the weaknesses of the search engine and the browser and something about the strength of paper.
- Understanding their *goals*. Every action has a reason. Understanding why people perform certain actions reveals an underlying structure to their work that they may not be aware of themselves. Although a goal may seem straightforward ("I need to find and print a TV listing for tonight"), the reasons behind it may reveal a lot about the system people are using ("I want to watch *The Simpsons* at eastern time so that I can go bowling when it's on normally, but I don't know which satellite it's going to be on and there's no way to search the on-screen guide like you can search the Web site").
 - Understanding their *values*. People's value systems are part of their mental model. We're often driven by our social and cultural contexts as much as by our rational decisions. What is the context for the use of these tools? Are they chosen solely because of their functionality, or do other factors apply? Is the brand important? (If so, why? What qualities do they associate with it? Are they aware of other brands that do the same thing?) Do they work in the context of others who use the same tools? If so, why does the group do things the way it does?

There are several ways to do a full analysis of the data you collect. Beyer and Holtzblatt recommend what they call the *affinity diagram* method (which is loosely patterned on the technique of the same name pioneered by Jiro Kawakita and known in the industrial quality management world as a *KJ diagram*). This method creates a hierarchy of all the observations, clustering them into trends. A paraphrased version of their method is as follows (see Chapter 9 of *Contextual Design* for a complete description of their method):

1. Watch the observation videotapes to create 50–100 notes from each 2-hour interview (longer interviews will produce more notes, though probably not in proportion to the length

of the interview). Notes are singular observations about tools, sequences, interactions, mental models—anything. Number the notes and identify the user whose behavior inspired it (they recommend using numbers rather than names: U1, U2, U3, etc.). Randomize them.

2. Get a group of people together in a room with a blank wall, a large window, or a big whiteboard. Beyer and Holtzblatt recommend 1 person per 100 notes. Preferably, these are members of the development team. By making the development team do the analysis, a group understanding of the customer's needs is formed, consensus is built, and everyone is up to speed at the same time. Have them block out the whole day for the work. The whole affinity analysis should be done in a single day.
3. Divide the group into pairs of analysts. Give each pair an equal number of notes (ideally, each pair should have 100–200 notes).
4. Write a note on a yellow Post-it (yes, *yellow*; Beyer and Holtzblatt are very specific about Post-it colors) and put it on the wall/window/board.
5. Tell the groups to put notes that relate to that note around it one at a time. It doesn't matter how the notes relate, just as long as the group feels they relate.
6. If no more notes relate to a given note cluster, put a blue note next to the group. Write a label on the blue note, summarizing and naming the cluster. They recommend avoiding technical terminology in the labels and using simple phrasing to state "the work issue that holds all the individual notes together."
7. Repeat the process with the other notes, labeling groups in blue as they occur.
8. Try to keep the number of yellow notes per blue group between two and four. One note cannot be a group, and often groups of more than four notes can be broken into smaller clusters. However, there's no upper bound on how many notes may be in a group if there's no obvious way to break it up.
9. As the groups accumulate, they recommend using pink notes to label groups of blue notes, and green notes to label groups of pink notes.

Eventually, you run out of yellow notes, and the group of analysts reaches a consensus about which notes belong in which group and how to label the blue, pink, and green notes (Figure 8.1). At that point, you have a hierarchical diagram that shows, to quote Beyer and Holtzblatt, “every issue in the work and everything about it the team has learned so far, all tied to real instances.”

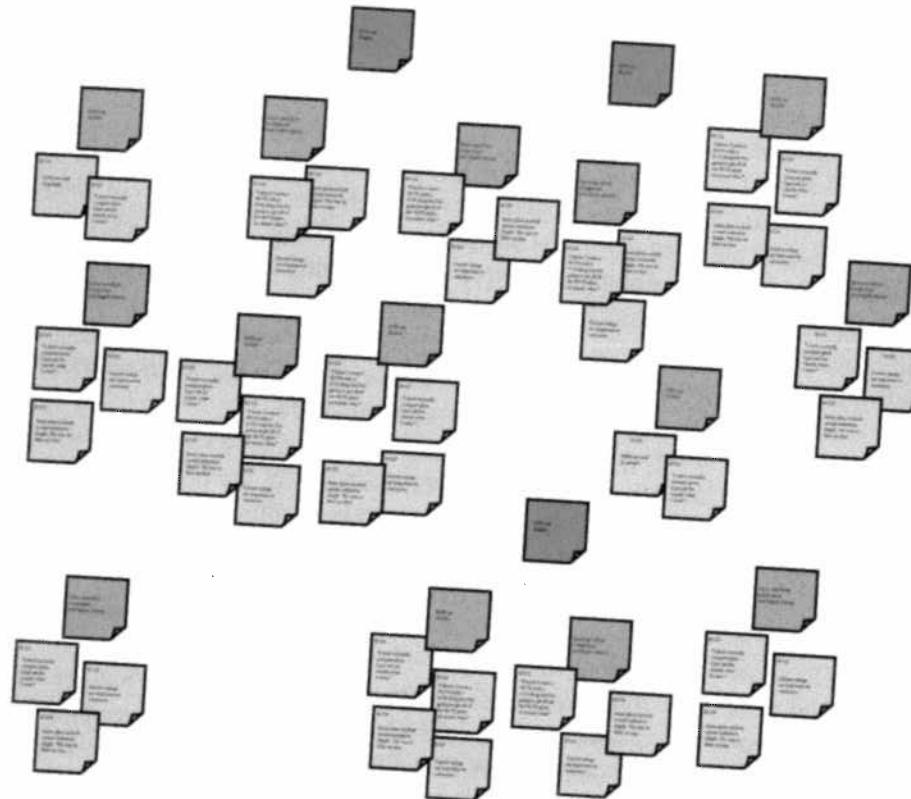


Figure 8.1 A portion of an affinity diagram.

The insurance broker observation might produce an affinity diagram with the following fragments:

RFPs are tools that collect most of the information

RFPs are flexible

RFPs are read by people. (note 35, U2)

“I don’t normally compare plans. I just ask for exactly what I want” (note 20, U1)

Plan comparisons provide critical information

Query specificity is important (and largely absent)

“I know I want a 90/70 with a 5/10 drug, but I’m going to get all the 90/70 plans no matter what” (note 55, U2)

Some plans exclude certain industries (legal). No way to filter on that. (note 43, U3)

Carrier ratings are important to customers. (note 74, U3)

Query output requires a lot of filtering

“Most of my time is fishing for information.” (note 26, U1)

In addition to the affinity diagram method, it’s also possible to do a more traditional kind of analysis based on an expert reading of the data. The observers can all meet and discuss their observations

and hypothesize about how those observations are linked and what they mean. Although not as rigorous, this process can cut through voluminous data (which the Post-it process often produces) and create workable models that adequately describe people's behavior and attitudes. It's also a process that can be done by a lone researcher if the development team is unavailable. However, the affinity diagram process is recommended, when possible, because of the benefits it provides in communicating people's needs to the development team and getting maximum use out of the collected data.

Building Models

In many situations, this will be enough to begin building models of where the unmet needs are and how solutions can be made to fit with people's existing work practices. Beyer and Holtzblatt define five models that can be extracted from the information in the affinity diagram: flow models, sequence models, artifact models, physical models, and cultural models.

- *Flow models* are representations of “the communication between people to get work done.” They show what information, knowledge, and artifacts get passed among the members of the development team. Their elements can be formal or informal, written or verbal. They seek to capture the interaction, strategy, roles, and informal structures within the communication that happens in a product’s life cycle.
- *Sequence models* represent “the steps by which work is done, the triggers that kick off a set of steps, and the intents that are being accomplished.” They show the order that things are done, what causes (“triggers”) certain steps, the purpose (“intent”) of each step, and how they depend on each other. They are sufficiently detailed to allow the team to understand, step by step, how a task is accomplished.
- *Artifact models* represent how people use real-world tools to accomplish their goals. Starting with a simple photograph, drawing, or photocopy of the object, artifact models “extend the information on the artifact to show structure, strategy, and intent.” They provide insight into the tools people use, how

Note Frequency does not equate to importance. Just because something happens often doesn't mean that it's more important to the design of a system than something that happens rarely. For example, most people observed may keep paper notes of a certain transaction, and they may do it several times a day. Although this is a prevalent problem, it may not be as important as the half hour they spend browsing through new documents because the search engine isn't up to date. The paper notes may get used often and are clearly compensating for a deficiency in the product, but maybe they represent a solution that's tolerable for the users. This makes them a less important issue than the search system, which is a problem for which they have no good solution at all.

they use them, what problems they have, and most important, why they're necessary.

- *Physical models* represent the actual physical environment that users do their work in. They provide an understanding of the layout of the workspace, the artifacts in the workspace, what controls people have (and don't have) over their environment, and how they use their environments to get work done.
- *Cultural models* represent an understanding of the users' values and how they see themselves. It places the product in the context of the users' lives and the real-world environment in which they live. It includes both the formal organization of their experience—their other responsibilities, the competitive climate—and the informal—the emotions associated with the work, the work environment, the users' esthetic values and style, and so on.

Producing Results

In many cases, the “Aha!” moments come either during the actual observation or in the affinity diagram creation phase. The aha! may be sufficient to turn the conceptualization of a product around and provide enough information so that product-specific (as opposed to problem-specific) research such as focus groups and paper prototyping can begin.

It's never good to jump to conclusions, but even if time is scarce, the results of the research should be consolidated into a document that sets out your understanding of the audience and compares it to the assumptions and scenarios you had at the beginning. Consisting of your thoughts about your users' mental models, tools, terminology, and goals, this document can serve as a “statement of understanding” that drives other research and feeds into the research goals as a whole. Return to it after every additional piece of user information is obtained, and attempt to understand the new information in light of the statement. If the new information is contradictory or confusing, it may be time to repeat the research with a new group of people.

If you have more time and resources to look at the information you've collected, do so. The data can provide a rich and subtle understanding of the mental models and task flows people use in

doing their work. Beyer and Holtzblatt spend a good deal of their book discussing just how to do this.

Task Analysis

The primary purpose of task analysis is to compare the demands of the system on the operator with the capabilities of the operator, and if necessary, to alter those demands, thereby reducing error and achieving successful performance.

—B. Kirwan and L.K. Ainsworth in *A Guide to Task Analysis*

There are a number of techniques that fall under the definition of task analysis. They range from structured, standardized questionnaires to unstructured interviews with usability experts. For this book, I use the definition that I believe to be the most common, that task analysis is a structured method of hierarchically analyzing a single task in order to understand the interaction of its components.

This means that it's quite closely associated with contextual inquiry, and in fact, the data for it can be collected during a contextual inquiry process. It differs from that procedure in its degree of focus. While contextual inquiry attempts to understand the entire context that surrounds and informs a task, task analysis focuses on the task itself. What is the exact order of things? What are the tools involved? Where is there flexibility in the process? What kinds of information do people need and use at various parts of the task?

When Task Analysis Is Appropriate

Task analysis is best used when you already know what problem you're trying to solve and you want to know how people are solving it right now. This generally falls in the examination phase of a spiral development process or the requirements-gathering phase of a waterfall process. It requires that you know what the task is and, roughly, who your audience is. Although it *can* be done when there's already a solution of some sort under consideration, it should ideally be done before effort has been invested in features and technologies since it's likely to induce a lot of revision in the fundamental assumptions of how solutions should be implemented.

So if you've already decided that you're going to be selling supermarket-style groceries online, but you haven't yet built the ordering system, now would be a good time to find out how a grocery shopper picks out groceries. Again, you'll be looking for many of the same things as contextual inquiry: tools, sequences, organizations, interactions. But whereas that process investigates the decision-making process people go through in determining what they're going to eat, how they're going to get it, where they're going to get it, and so forth, task analysis will concentrate solely on the task of buying groceries in a supermarket.

In addition, although it can reveal places for interaction with advertising or promotion, the goal-oriented nature of the analysis makes the process better at uncovering opportunities for improving usability rather than promotion or advertising (Table 8.2). That said, there's still much untapped potential in the technique that can produce interesting results when applied to the domain of the entire user experience rather than just their immediate pragmatic goals.

Timing	Activity
$t - 2$ weeks	Organize and schedule participants.
t	Begin interviews.
$t + 1$ week	Complete interviews. Review videotapes and take notes.
$t + 2$ weeks	Begin results analysis.
$t + 4$ weeks	Complete results analysis. Present to development team. Distribute documentation.

How to Do Task Analysis

The form of task analysis described here is a somewhat relaxed combination of the two methods described as *decomposition methods* and *hierarchical task analysis (HTA)* in Kirwan and Ainsworth's *A Guide to Task Analysis*.

Preparation

The preparation for task analysis is almost identical to that of contextual inquiry. You need to pick a target audience, recruit them, and schedule them. You need to plan for the interview by learning the domain and to develop scenarios about how you feel the task will commence. Finally, you need to be organized in the way that you conduct the interview so that you have enough supplies, the right equipment, and the knowledge to use it.

Gathering the Information

The work published in Kirwan and Ainsworth's book was done before the advent of the Web, and it concentrates on installed software and the issues surrounding it. So although they discuss starting with published information and experts as sources for uncovering the actions that make up a task flow, you won't often find such sources for most Web-related work. Moreover, what's important is not how people are supposed to do their work, but how they actually do it. In Kirwan and Ainsworth's task domain, nuclear power and nuclear chemistry, safety requires that people follow the stated rules much more rigorously than for just about any other field. In modern software environments, direct observation is again the basis for data collection.

The interview format is much the same as contextual inquiry: the interviewer acts as an apprentice to the person performing the task, watching as the participant does the task and explains key elements. The apprentice watches and attempts to understand the nuances of the task, eliciting occasional explanations of motivations and options.

Where it differs from contextual inquiry is in terms of focus. Everything should be geared toward understanding how the participant performs the task at hand.

- What do they see as the options at any given point?
- What are the tools available?
- How do they choose one over the other?
- Where do they change their minds?
- How variable is the process?

- Where do they make mistakes? What are common mistakes?
- What causes the mistakes, and how are they corrected?
- What are the inputs to the process? The outputs?
- What is the frequency and importance of the tasks they perform?
- What are the risks of failure?

The kinds of information gathered should be similarly structured and focused as contextual inquiry, but with extra emphasis on sequence information and the exact tools used.

Here's a snippet of the notes from a session observing an interior designer picking out a couch (for an online furniture-buying tool).

"I get out my catalogs. I go to [manufacturer A] first, unless it's for a waiting room, then I go to [manufacturer B]. I have a big shelf of catalogs. Recently, I've been getting on the Internet, but I get better results on the catalogs. Hard to find things. Each Web site is so different, it takes time to find your way through the Web site."

Gets four catalogs: A, B, C, D. Gets the [B] catalog. Looks in index for "couch," flips to couch section.

"I usually find a couple of options and weigh cost, availability, delivery, if there's something similar I've seen. You can't tell color from pictures, and they're often slow about getting fabric swatches. I talk to sales reps on the phone and try to get as complete a picture as I can. Some of the companies we don't do so much business with will come in and also try to sell me other stuff."

Marks several sofas with small notepads, in preparation to call the manufacturer about fabric/color availability.

"I know I can get this one pretty fast since I've gotten a couple before, but I'm not sure if they can offer it in the fabric we want. I'll make a note to ask about it. Sometimes they can make it out of a different fabric even if they don't officially offer it in that, but it depends."

(Continued)

"I total it all up. This is what we need, and this is what it will cost, and this is how long it'll take to get it made, assuming they have the fabric in stock. I'll fill out purchase requisitions, going through the catalog and picking out the options I want. I put together a description for the client. Project explanation, list of things to buy, schedules, and so on. I may not have time to explain it all to the client, so I have to make sure the write-up is clear enough."

"Different vendors offer different warranties. I don't shop for warranties, though, I shop for matches and manufacturing time. Cost is important, too, but once the client has signed off on a budget, that limits your options right there."

Prepares to call vendor.

How to Analyze Results

Task decomposition and *hierarchical task analysis* are complementary techniques. One describes the inputs, outputs, and triggers for the actions in a task while the other arranges the actions into a coherent flow. The order in which these two techniques are used depends on the primary needs. If you're mostly interested in how the components of a task fit together (for example, if you're looking for where a tool can optimize a task's speed), then start with HTA and flesh out key details with decomposition. If you're interested in making a tool that fits with existing practice, then start with decomposition and put the pieces back together in a hierarchical form.

Break the Task Apart (Task Decomposition)

Task decomposition is the process of breaking the task into its component actions. This is easier said than done. If someone is operating a control panel with a bunch of buttons and dials, then every time he or she presses a button or reads a dial, it's an action. But when that same person is shopping for a new car, the decomposition is more difficult since the steps taken—the choices, the capitulations, the comparisons—are not nearly as obvious. In such situa-

tions, encourage the participant to speak all of his or her thoughts aloud and prompt him or her for explanations of specific actions.

Keep the end goal in mind since this will help you pick out the most relevant aspects of each action. If you have a specific tool or solution in mind, you can emphasize one or the other in your decomposition. Sometimes you can work backward from the solution you have in mind to the problem people are facing to see if one matches the other.

Describe each action. Kirwan and Ainsworth list several dozen categories that have been used to describe various facets of an action. Some common ones include

- *Purpose*. Why is this action here? How does it move the task toward its goal?
- *Cues*. What tells the person that it's time to perform this action?
- *Objects*. What does the action operate on?
- *Method*. What is the action?
- *Options*. What other actions are available at this point? How did this action get chosen?

While you're describing the actions, do some error projection. Ask what would happen if the action weren't done or if there were an error. If there are enough actions with interesting error conditions, you can even make errors a separate category you use to describe each action.

For each action, provide answers in as many of these categories as you can. Often the process of describing an action will create questions and inspire new interpretation, so it can be useful to walk through the decomposition a couple of times to make sure that it's consistent and thorough.

Here is a snippet of the furniture-buying decomposition.

Prepare for decomposition by determining the categories that are likely going to be useful ahead of time and then creating forms that you fill out as you observe. That way, you know what information you have about each action and what else you need. In situations where there are a lot of actions in every task or they come quickly, this can be a pretty daunting method of working. In such situations, it may be better to observe the task closely and decompose it into its components later.

Action Name	Purpose	Cues	Objects	Method	Options
List requirements	Clarify choice categories		Word template: size (h,w), color, budget, style notes	Talk to client	
Get available catalogs	List available options	Catalogs, requirements list	Compare options in catalogs with requirements		
Set order for catalog perusal	Start with best-known/best-chance manufacturers	Knowledge of vendor's options	Catalogs, requirements list	Flip through catalogs, comparing options with requirements	Go to A first, unless it's a waiting room, then B
Mark items in catalog	Get primary candidates		Catalogs, requirements list	Visual inspection and comparison to list	
Mark items that need follow-up	Separate items for further investigation	When it's not clear whether all options are available for a specific item	Catalogs, marked items	Visual inspection and comparison to list	
Investigate marked items	Complete list of available options based on requirements	List of items for further investigation	Catalogs, list of items needing follow-up	Call reps	
Total options	Make final list for client	All follow-up is completed	Completed options template, budget template, requirements list	Fill out budget template with options and costs	

Guerilla Task Decomposition

In many situations, it's not practical to do this level of task decomposition because of resource or time constraints. In such situations, you can use a more informal version of this process. Use the basic ideas of task decomposition, but instead of going all the way to the atomic task and filling out all the facets, concentrate on larger ideas and use only the facets that you think are most important. If after several interviews, you feel that you can write down the basic tasks, do so on Post-its, create some preliminary organization, and flow with them. Then, as you interview more people, you can adjust these to fit your new information. This is not as thorough of a process, but it can be a lot faster.

Put It Back Together (Hierarchical Task Analysis)

HTA consists of taking the individual actions that make up a task, putting them in a hierarchy, and creating rules for how to move through the hierarchy. The final product is a flowchart that details the sequence that leads up to every action, the choices that are made, and the consequences of actions.

HTA doesn't have to be done following task decomposition. It's possible to do hierarchical task analysis by just sitting down after having interviewed and observed users doing a task. The procedure is a top-down analysis method, which is essentially a formalization of the process that interaction developer, information architects, and technical writers go through when they stand up at a whiteboard and try to understand how something is done.

1. Start with a goal. This can be the person's ultimate goal ("Furnish the house"), but abstract end goals often require analysis that's too involved for straightforward task analysis. You may want to start with something more direct ("Get list of available couches") and work down from there.
2. Determine the subgoals for that goal. Subgoals are all the things that have to happen before the main goal can be met. For example, if your goal is to get a list of available couches, you may need to get a list of couches from the catalogs and a list of couches that are not listed (or unavailable) from furniture company representatives. You should rarely have more than three or four subgoals for every goal. If you do, you may

want to create a couple of intermediate goals that will subsume several of the intermediate tasks.

3. Determine how those actions have to be arranged and create a plan for how they flow together. This includes determining which ones have to follow which others, which are optional, which can be swapped out for each other, and so forth. When goals don't have a natural flow, pick one that seems to work and adjust it later if it turns out to be awkward.
4. Repeat the decomposition with each goal and subgoal until you are down to individual actions.

The end result is a diagram in the familiar “boxes and arrows” style, where a single goal has been turned into a goal tree, with all the subcomponents of every goal beneath it and linked to it, creating a procedure for accomplishing that goal (Figure 8.2).

When you've formally decomposed a task, the procedure is nearly identical, except rather than making up the list of goals and relationships on the fly, you use the ones that you've extracted from your interviews. The procedure to do this is a bottom-up method similar to the way an affinity diagram is created in contextual inquiry.

1. Start by ordering and clustering actions into groups that are all associated with the same result. Actions that occur in parallel are optional or are alternatives for each other and should be labeled as such.
2. Label the clusters with the results. These are the subgoals.
3. Order and cluster the subgoals in terms of what ends they achieve.
4. Label these clusters. These are the goals.
5. Repeat until all the actions have been clustered and all the clusters have been ordered and labeled.
6. Create a diagram by organizing the subgoal clusters underneath the goal clusters and linking these with arrows that show how goals are related to their subgoals.

This procedure produces diagrams identical to Figure 8.2.

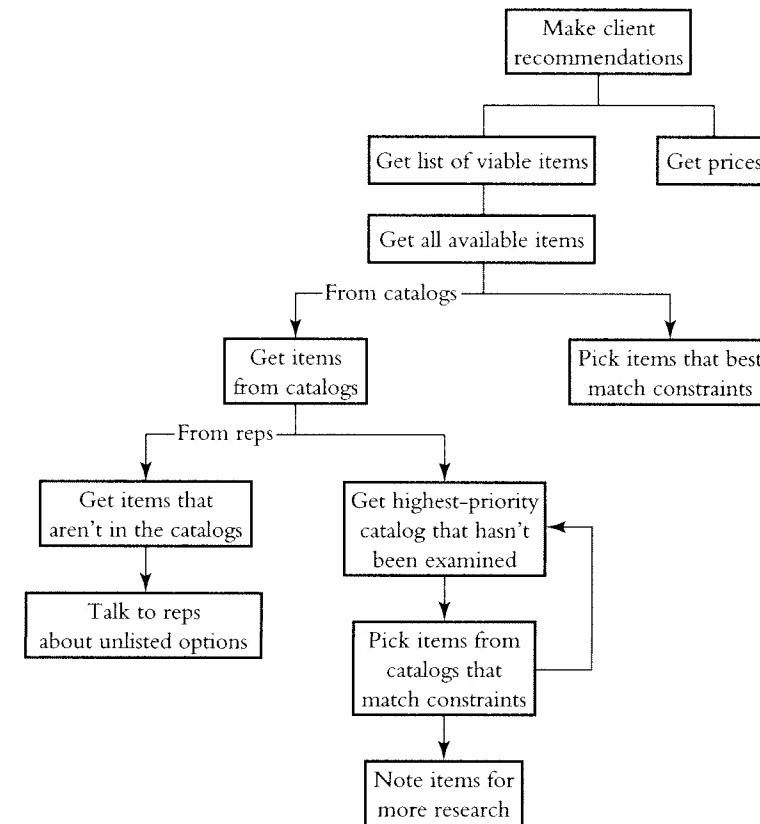


Figure 8.2 Couch-buying analysis diagram (segment).

Checking for Diagram Validity

Once you have your diagram, you can check it for validity and completeness using a technique mutated from the *function analysis system technique* (FAST, also described in Kirwan and Ainsworth).

- The boxes in the chain above a given box in the hierarchy answer the question “Why do I _____?” where _____ is what's written in the box. So using the example, when asking “Why do I talk to reps about unlisted options?” the diagram answers “Because you need to know what items are available that aren't in the catalog so that you can make a complete list of all available items to make a list of viable options for the client recommendation.”

(Continued)

- Boxes below a given box answer the question "How do I _____?" So asking the question "How do I get items from a catalog?" is answered by "Find the highest-priority catalog that hasn't been examined and pick items from it while noting down which ones require more research."
- Items with arrows only go in one direction (generally "how"). If you can answer both of the above questions for every box, your diagram is complete.

When you've completed a task analysis diagram, walk through it with a few people who know the task to make sure that it matches their experience and that the sequences, tasks, and goals are familiar to them from their experience.

Granted, this particular method can be time consuming and may be unable to capture the subtlety of certain tasks, but it forces deep consideration of the task that can reveal the true nature of underlying processes and can disentangle complex interactions. It's much easier to create a product that works the way that people do when you've understood how they behave. Task analysis provides a specific framework for matching people's existing behavior and satisfying their immediate needs. Moreover, it can demonstrate weaknesses and redundancies in the product since it had been described before the analysis was done: there may be tasks people are trying to accomplish that the product was not originally envisioned to assist in, or there may be inefficiencies in the way that people currently perform their tasks that the product can streamline.

The task analysis diagram, if accurate, allows developers to have a consistent road map and a thorough checklist of everything that the product needs to do (the list of goals) and the order in which it needs to be done (the task flow).

Card Sorting

Card sorting is a technique that's used to uncover how people organize information and how they categorize and relate concepts.

When Card Sorting Should Be Done

Card sorting is best done when you know what kind of information needs to be organized, but before an organizational solution has

been implemented. It is usually done after the product purpose, the audience, and the features have been established, but before an information architecture or design has been developed, putting it somewhere in the middle of the design process. This differentiates it from contextual inquiry and task analysis, which both come at the beginning of the development process.

In addition, since it's a fast and easy technique, it can be used whenever you change your information structure or want to add elements to an existing structure.

The Card Sorting Process

Preparation

The goal is to get perspective on how your intended audience understands your proposed information space. So unlike other techniques, there aren't any additional constraints on the kinds of people you should invite, other than recruiting people who fit into your target audience. Typically, recruiting between four and ten people from that audience will give you a good perspective on organizing the information.

Schedule people one at a time so that the participants don't feel pressured to compete. You can schedule several people simultaneously if you have facilities where they can sit quietly and if there is someone nearby who can answer questions (though if you have only one person as a monitor, stagger the schedules about every 15 minutes so that the monitor has time to give each participant an introduction to the technique). An hour is more than sufficient for most card sorting studies.

Getting the Cards Together

The core of the card sorting process is, not surprisingly, the cards. On a deck of sturdy identical note cards, write the names of the things that you want to organize (or use a word processor mail merge and print on mailing labels that you then stick to index cards). These can be the names of specific sections, terms you're considering using, concepts behind the various sections of your site, images that you want to use, or even descriptions of individual pages. Use identical cards, except for the text, to minimize distraction.

You can have as few or as many cards as you want, though the size of a standard card deck (52) strikes a good balance between not providing enough cards to make adequate categories and providing so many that it's overwhelming. If you have hundreds of categories that you would like to try this technique on, consider breaking them up into more manageable chunks and doing multiple sets of tests.

The words on the cards should reflect what you're trying to test. If you're trying to uncover how people organize concepts, explain the concepts on the cards with a sentence or two. However, if you're trying to see how people understand a set of titles without necessarily knowing your definitions for them, you can just write the titles on the cards.

The Sort!

After bringing the participant in and going through all the initial formalities, introduce him or her to the concept. Say something along the lines of this.

This is a stack of cards. Each card has something that you might see on the Web site. I'd like you to organize them into groups that make sense to you. Take as much time as you need. There are no right or wrong groupings. Try to organize all the cards, but not everything needs to belong in a group. You won't have to provide a reason why cards belong in the same group, so if a group feels right, go with it. Focus on what makes sense to you, not what may make sense to anyone else.

Provide a stack of Post-it notes, several pens, and a pile of small binder clips or rubber bands. After they're done grouping, ask them to label the groups if they can, but not every group necessarily needs a label (don't tell them that they'll be labeling ahead of time since that tends to bias people to organize based on labels rather than on what they feel are natural groupings). When they're done, ask them to clip or rubber-band the cards and place the label on the groupings.

Finally, ask them to organize the groupings into larger groups without moving any cards from one group to another, naming the

larger groups, if any names come to mind. Then, with larger binder clips or rubber bands, bind the metagroups together and wrap up the session.

Card Sort Analysis

There are two ways to analyze the output from card sorting, a formal way and an informal way.

The Informal Way

When you have the clusters from all the participants, look at them. Copy the clusters to a whiteboard. By eyeballing the trends in the clusters, you can infer how people intuitively understand the relationships between the various elements. For example, if people put "News," "About us," and "What we like" together, it tells you they're interested in putting all the information coming from your perspective into a single place. However, if they group "News" with "Latest Deals" and "Holiday Gift Guide," then maybe they associate all the information that's timely together. The difference between these two can mean the difference between an information architecture that matches its users' expectations and one that forces them to hunt for information.

You can look at the clusters as a whole or follow one card at a time through the participants' groups to see the kinds of groupings it's put into. Don't treat the clusters literally. People's existing organizations may not make a scalable or functional architecture. Instead, look at them for the underlying themes that tie them together. Pay attention to the cards that people didn't categorize or that were categorized differently by everyone. What about the card is giving people trouble? Is it the name? Is it the underlying concept? Is it the relationship to other elements?

When you've gone through all the clusters produced by all the participants and listed all the themes, go through the names they've assigned to the groups. Looking at the labeling and the labels' relationships to the clusters underneath, you should have the basis for creating an architecture that's close to how your user base expects the information to be organized (and even if it's not used by the information architect, the terminology can be useful to marketing when they're explaining the product to potential clients).

The Formal Way

Cluster analysis is a branch of statistics that measures the “distance” between items in a multivariate environment and attempts to find groupings that are close together in the variable space. This is exactly what a whiteboard-based card sorting analysis does, but it’s a mathematical way to do it thoroughly and consistently. It allows you to uncover groups of objects that are similar across many dimensions, but may not be obviously alike in any one of those dimensions. Since people have trouble visualizing things in more than three dimensions, and there are often more than three variables that can determine similarity, the technique is used to “see” clusters that would have otherwise gone undiscovered.

The classic example of cluster analysis is from marketing research, and it’s the grouping of target markets. People are different in many ways: their age, their income, their race, where they live, their gender, what they buy, and so on. Picking out groups along any one of these axes is relatively straightforward, but picking out clusters along all of them is much more difficult. How do you define a yuppie? Do they all drive Land Rovers? No, some drive BMWs, and some people who drive Land Rovers aren’t yuppies. Do they all make \$120K per year? No, some may make \$30K and live as if they made \$120K. Figuring out the clusters in society is hard, but cluster analysis can often extract useful distinctions.

In terms of card sorting, it works in reverse. It’s used to find the underlying variables by looking at the clusters people make. Are certain things grouped together more often than other things? Are there hidden relationships between certain cards? These are all things that are hard to see by just looking at the cards.

Unfortunately, the mathematics of cluster analysis is nontrivial and can’t be easily done without a computer. Statistical packages such as SAS and Statistica contain modules that can do cluster analysis, but these are expensive and require an understanding of the statistical procedures used in the analysis.

Fortunately, there is a piece of software that is designed specifically to do cluster analysis for card sorting data and is, as of summer 2002, free. IBM’s User Involvement Group made EZSort available to the usability research community in 1999, and it’s made the process of analyzing card sorting much easier. The program takes as input the groups created by the participants and produces tree graphs that show the relationship between groups of cards, revealing

clusters of cards and clusters of clusters. These diagrams make it much easier to separate strong, common affinities from casual similarities and to see larger themes that would have been difficult to see through a haze of cards and groupings.

As of fall 2002, it’s available from

http://www-3.ibm.com/ibm/easy/eou_ext.nsf/Publish/410

The process of using it is straightforward: the names of all the cards and participants are entered into the program, then each person’s clusters are re-created in the software by dragging the card names into piles; the names for the clusters are associated with these virtual piles, and the software is run. It produces a diagram that represents the relationship between all the items using a tree diagram (Figure 8.3). The more distant one “branch” is from another, the less related they are.

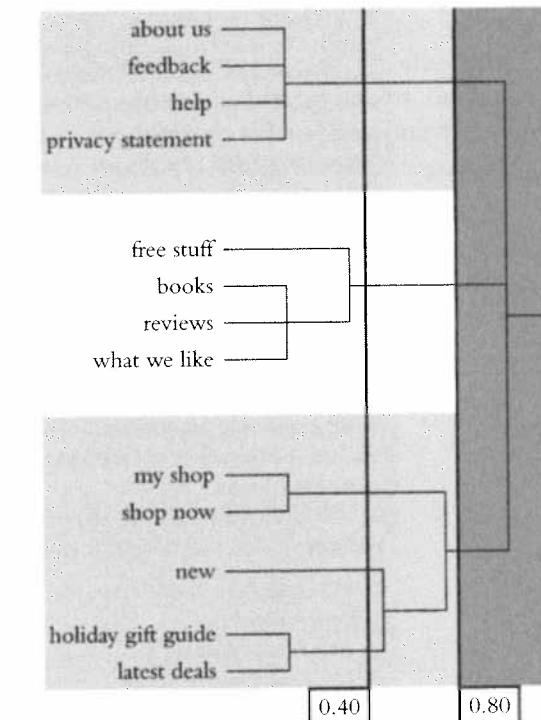


Figure 8.3 EZSort diagram for a typical shopping site.

Note For more information on EZSort, see the documentation or the paper by Jiamming Dong, Shirley Martin, and Paul Waldo that can also be found at [http://www-3.ibm.com/ibm/easy/eou_ext.nsf/Publish/410/\\$File/EZSortPaper.pdf](http://www-3.ibm.com/ibm/easy/eou_ext.nsf/Publish/410/$File/EZSortPaper.pdf)

This diagram shows that information about the company strongly clusters (the tree at the very top of the diagram), which likely means that people expect all that information to be in the same place (or at least treated in the same way). Likewise, general product information such as recommendation, reviews, and “free stuff” are all considered similar in some way, with “free stuff” being a bit less similar to the others in the same group. The other elements also form smaller clusters. The two vertical lines represent “threshold” for membership—how close items must be in order to be grouped together—with the number below each one representing the calculated affinity that the line represents.

Card Sorting to Prioritize

Although card sorting is primarily an organization or naming technique, variations of it can also be used for other purposes. Jesse James Garrett, an information architect and one of my business partners in Adaptive Path, and I developed a technique that uses card sorting to understand how people prioritize features.

A group of 12 participants were asked to organize 50 cards that had descriptions of current and potential features for presenting current events on Web sites (this was done when researching how people read news online, as a break during a long one-on-one interview). Each card had a single sentence that described a way that information could be presented or organized on a Web site. The participants were first asked to place the cards into one of four piles describing how *valuable* they felt the feature would be to them, as they used the Web right now. The four piles were titled “Most Valuable,” “Somewhat Valuable,” “Least Valuable,” and “Not Valuable.” As participants placed the cards in these piles, they were asked to narrate their thoughts about the topic. This was done to understand why they made certain choices and how they thought about the features.

After the participants completed organizing the cards, they were asked to repeat the exercise, but this time only using the cards that they had put into the “Most Valuable” pile and with a different set of criteria for the placement. The new task was for them to organize the features in terms of how *frequently* they felt they would use each of the features. This was done in order to differentiate between the features that attracted people to a site and those that they felt were the most immediately useful.

To understand where the participants’ values lay, each one of the categories was given a numerical rating indicating the strength of preference, from 0 to 5:

- 0 - Not valuable
- 1 - Least valuable
- 2 - Somewhat valuable
- 3 - Most valuable, rarely used
- 4 - Most valuable, sometimes used
- 5 - Most valuable, used often

The participants’ choices were then rated, and the *median value* of their ratings was calculated. Since a number of choices had the same median, it was necessary to further organize the list. The *standard deviation* of the choices was calculated and represented the degree of agreement among all the ratings for a given category, with lower deviations representing greater agreement. Both median and standard deviation are defined in Chapter 11. The list was thus ordered first by preference and then by agreement, which gave the development team a much clearer idea of their customers’ values and in turn helped them prioritize their own development efforts.

The card sorting process sheds light on people’s existing understanding and preference, and it can show subtle relationships that may not be obvious by just examining a list of clusters. It also provides an idea of how concepts relate to each other since what may seem like a strong relationship when casually examined may turn out to be weaker when actually analyzed.

Contextual inquiry, task analysis, and card sorting can take you a long way toward revealing how your target audience understands their environment and what their needs are. As prescriptive techniques, they can focus a project early on, eliminating many wrong turns and reducing the need to ask major, fundamental questions about the product later on, when development should concentrate on honing the product’s experience, not its purpose.