11-2005

Using Future Context in Personal Information Retrieval

Paul J. Timmins
Worcester Polytechnic Institute, ptimmins@cs.wpi.edu

Craig E. Willis
Worcester Polytechnic Institute, cew@cs.wpi.edu

WPI-CS-TR-05-17 November 2005

Using Future Context in Personal Information Retrieval

by

Paul J. Timmins
Craig E. Wills

Computer Science
Technical Report
Series

WORCESTER POLYTECHNIC INSTITUTE

Computer Science Department
100 Institute Road, Worcester, Massachusetts 01609-2280

Delivering Relevant and Useful Information with IMPACT

Paul J. Timmins and Craig E. Wills
Computer Science Department
Worcester Polytechnic Institute
Worcester, MA 01609
{ptimmins,cew}@cs.wpi.edu

Abstract

With nearly universal access to vast information sources, it is difficult to discern what is important from what is not. Users, especially mobile users, are faced with a tradeoff: either check many information sources frequently or miss important information that may impact them. Improvements in information retrieval are required to consolidate diverse information and reduce the overall effort required to monitor multiple sources. Effective personal information retrieval requires context awareness, and in this work we propose the use of knowledge of future context as the basis for identifying and prioritizing information. By taking into account the user's planned activities, it is possible to be continually informed of information that may have an impact on the user, with minimal disruption to the user.

This paper provides an overview of the issues surrounding the use of future context in information retrieval and presents a context-aware retrieval system built to explore these issues: the Integrated Management of a Personal Augmented Calendar Tool (IMPACT). IMPACT employs future context to prioritize, organize and deliver information, leveraging the user's calendar to determine future context. The goal of the IMPACT system is to provide the user with all information that will impact activities in a timely manner, while minimizing the effect of interruptions on the user.

1 Introduction

The Internet allows us to stay in-touch and informed in ways never before possible. Instant access to vast information resources, from news to email to weather, allows us to plan and react with relevant and timely information at our fingertips. Yet, our ability to stay in touch with these resources is limited. There are simply too many resources, with far too much information, to constantly monitor. Confounding this problem is the fact that often only a

portion of information from a given source is of importance, requiring users to filter out what is important and what is not. This results in information overload [18] for many people, a situation where the amount of information overcomes the ability to receive the information.

To prevent information overload, we reduce the amount and rate of information that we receive. Web users, for instance, reduce their information consumption by checking web sites at infrequent intervals. This approach leads to missed or delayed delivery of important information. Experienced users may subscribe to notification services, choosing sources that are likely to provide useful information, such as traffic along local highways. Browser extensions, such as ForecastFox [1], integrate information sources of interest to the user directly in a commonly used application, requiring little effort by the user. These services are useful if well-matched for relatively static information needs of a user, like local weather, but fail when the information needs for a user change, such as when traveling.

Other factors affecting personal information retrieval are current browsing and notification approaches, which are inefficient for a number of reasons. Browsing requires the user to choose a source and seek out useful information, requiring the user to focus their attention on the browsing. Notifications typically interrupt the user based on when the information becomes available, not based on when the user needs the information. RSS aggregation [24] can help address the problem of monitoring a large number of sources, but does not easily deal with the changing interests and context of the user. The root of the problem is that the user must sift through information to determine what is useful and what is not. Users are therefore faced with a tradeoff—spend more time filtering information from a number of sources or monitor a smaller number of sources and miss useful information.

Systems that rely on static rules and preferences, such as receiving traffic alerts for a particular city, cannot account for the fact that the interests and priorities of a user are dynamic. One key factor in determining the usefulness and urgency of information is whether it impacts an event or task the user will perform. From traveling to meetings to research to social events, what the user does determines a large part of what the user is interested in. For instance, weather in another city and traffic along a highway may impact travel plans. Emails on a certain subject or from particular people may impact plans for meeting and social activities. The user's future context can therefore be used to prioritize information based on how significantly it impacts the individual activities, reducing the time and attention required to filter through a large number of information sources.

To determine whether an event will impact a user's future activities requires knowledge about the user's future context. Where the user will be, what the user will be doing, and with whom, all are important factors influencing the importance of information. In this work, the user's calendar is used as the source for the user plan—what the user intends to do. Knowledge of future context allows the incorporation of dynamic interests and priorities in the selection and filtering of information for the user. For example, when a user is traveling between locations, they are likely to be impacted by the local weather and traffic. The user plan also provides a mechanism for determining when interruptions are least disruptive, such as during breaks between activities. For information that does not require the user's immediate location, such as driving directions, information can be provided in the context of the user's activities, allowing the user to quickly find information that relates to an event, again reducing the time and attention required to find a particular piece of information.

The problem of timely notifications is a particular challenge for users away from their

desktop or laptop computer. Mobile users have limited access to information, despite the proliferation of Internet-enabled mobile phones and PDAs, primarily because mobile users are focused on activities other than a computing device. Aggravating this problem is dominance of the browsing-oriented model, which requires the user's focused attention on retrieving information. With the limited capabilities of mobile devices, compared to desktops and laptops, browsing is that much more difficult. Today's model of browsing is simply the wrong, or at least incomplete, approach for delivering important information to mobile users. Notifications, such as text messages, interrupt the user to inform them of new information. However, any interruption of the user for non-important information is undesirable, requiring careful consideration of how and when information is important.

This paper describes our project to design and build a context-aware computing system, the Integrated Management of a Personal Augmented Calendar Tool (IMPACT), which uses a user plan to prioritize, organize and deliver information. Its goal is to inform the user of any and all information that will impact the activities that a user intends to perform. To accomplish these tasks, the user's plan, which contains the user's current and future context, is built up from static preferences and calendar-based data. For each event in the plan, information sources are queried and assessed according to rules. Information is assessed for level of impact (severity) and urgency. The urgency is used to prioritize information, which is delivered to the user at times and via means designed to minimize disruption.

The IMPACT system shows how useful relationships between events and commonly available information sources, such as web sites and emails, can be built using the limited context available from calendars: date ranges, keywords, and locations. This work is novel in that it solely uses the user's future context for determining relevance of information, providing automated retrieval and delivery of information. Further, this work presents a context-aware application that brings together the concept of future context with the distinct areas of HTML extraction and interruption management. The importance of this work follows from the rapid growth of the World Wide Web, which has forced users to make a tradeoff between high rates of interruptions and the risk of missing potentially critical information.

The remainder of the paper is organized as follows. Section 2 explores the concepts and considerations for the design of a context-aware retrieval system. Section 3 outlines the IMPACT system, demonstrating a framework for personal information retrieval. Section 4 discusses how a system such as IMPACT should be evaluated. Related work is discussed in Section 5. Section 6 provides a summary and plans for future work.

2 Context-Aware Information Retrieval

An effective personal information retrieval system will improve a user's awareness of important information with minimal user effort and disruption. The goal of such a system is to bring information together from sources of interest, aggregating a number of sources into a single interface. RSS aggregators, for example, consolidates headlines and topics into a single interface [14], allowing the user to be informed when new information is available. Personal information retrieval systems reduce the need for the user to check individual sources, such as web sites, for new content. As a result, personal information retrieval reduces the time it takes for the user to be informed of new information, allowing the user to focus on their

task at hand, not on the obtaining and checking information sources.

It is not easy to determine what is of importance to a user. User interests change. What is interesting and relevant to a user during a weekday is often different than during the weekend. Where the user is, where the user is going, and what the user will be doing are all significant factors influencing user interest. Therefore, any personal information retrieval system that is not aware of the user's context, that relies on static rules to determine interest, must either provide a superset or subset of important information leading to either excessive disruption or poor awareness.

Using standard concepts in the area of information retrieval [29], the following criteria are the primary objectives of a personal information retrieval system.

- Recall: Is all important information delivered to the user?
- Precision: Is only important information delivered? Is the information properly prioritized and filtered?
- Minimal Disruption: Is the user unnecessarily interrupted?

Recall directly affects the systems ability to improve the user's awareness, with greater recall comes higher awareness. Poor precision means that some information is delivered to the user that is not important. To be minimally disruptive, the system must take into account the disruption caused by delivery of information, a factor which also depends on the user's context.

These issues of recall, precision and disruption depend on the determination of importance, a subjective concept. In this work, importance is defined as that information which is likely to affect the users behavior and impacts the user's activities. Impacting information is that which affects the users current or future activities, such as inclement weather affecting travel or an urgent email relating to a upcoming meeting. It is this class of importance, impact, that is the focus of this research.

Determining what will impact the user requires knowledge of the user's context in the future, what the user will be doing. With knowledge of the user's future context, it is possible to aggregate information within the user's context. The conceptual architecture of a system for information retrieval is outlined in Figure 1. This figure shows the components of user plan generation, context-aware retrieval, impact assessment and information delivery. Such a system is intended to meet the goals of recall, precision, and minimized disruption as discussed in the remainder of this section.

2.1 User Plan Generation

The user plan is an approximation of the user's context over time, including where and what the user will be doing. It provides a basis for determining what will impact the user, what information the user is likely to reference, and when interruptions will not significantly disrupt the user. In determining a user plan, multiple sources of context can be considered as potential predictors of the user's activities, as shown in Figure 1. A well-maintained electronic calendar is likely the best basis for the user's future context, as it explicitly reflects the user's intent and interests. To do lists may indicate those tasks which the user is likely to perform in the near future, although the exact timing is not known.

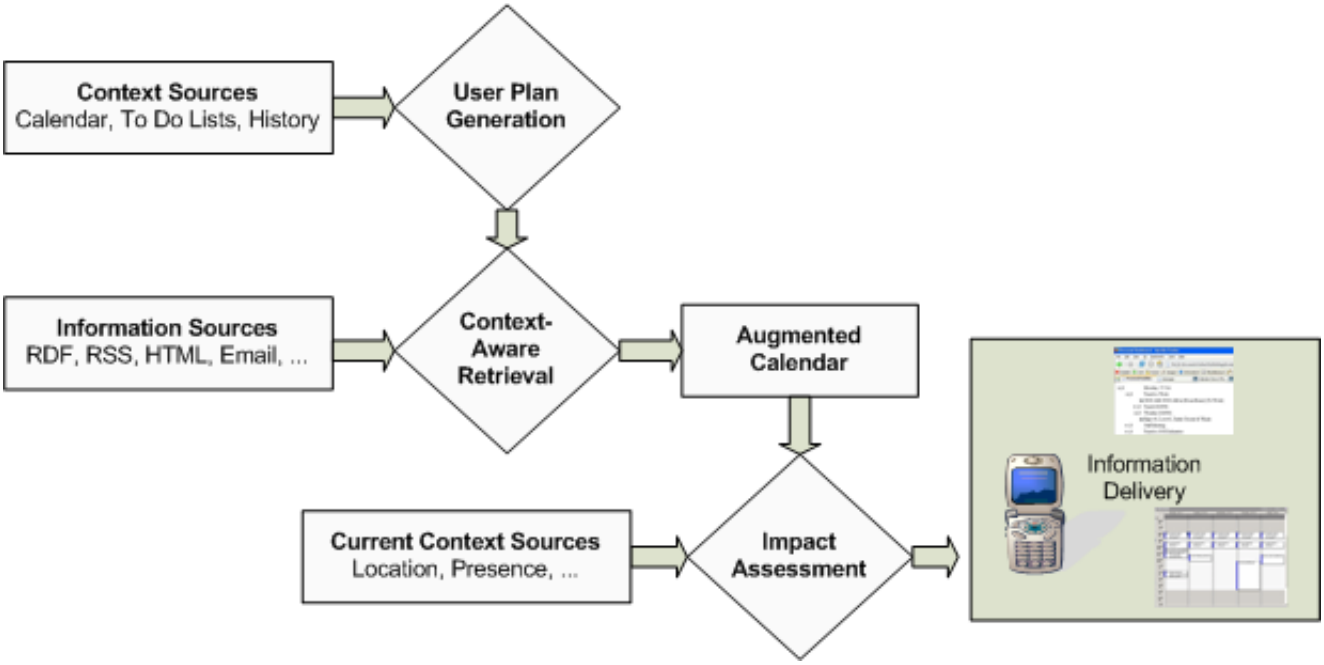

Figure 1: Conceptual Architecture.

Calendars record where and what the user will be doing, as well as other events of importance to the user. Events in a calendar range from business meetings to lunch plans to doctors appointments to daily commutes to vacation travel. Calendar events may also be used for reminders and information about other people's activities, such as a business partner's travel plans. The purpose of the calendar is to provide a list of events summarizing what the user either intends to do, or wishes to be reminded of, along with date, time, duration, recurrence, and often important details such as the subject, attendees and location. Even calendar events that user will not participate in may be of importance, as they reflect events of significant to the user.

Another source of future context are to do lists. To do lists provide a secondary source of unscheduled, but still planned, tasks that the user intends to perform at some point in time. Tasks in to do lists have broader time frames than scheduled activities, such as the user intending to complete a given task by the end of a week. Information that affects the tasks can impact the user and be of importance.

The user's history also can be used to predict future behavior [26]. Prediction of future context may be best integrated into the calendar information to infer behavior from patterns, and improve the accurate of the user plan. For example, if the user leaves work early every Friday, then the user plan should reflect this as a typical behavior.

2.2 Context-Aware Retrieval

The objective of context-aware retrieval is providing the highest recall of information; that all related information for a given event can be identified. Figure 1 shows the role of context-aware retrieval, receiving inputs from the user plan and information sources, and augmenting

the user plan with related information for each event. Retrieval can occur once a user plan has been constructed, and related information is extracted for each event to provide a single, aggregated source of information for the user. The related information provides the basis for determining whether and how to notify the user of specific information items. Information is then hierarchically associated with each event. For instance, weather would be a part of each event, with attributes of temperature, humidity and forecast.

Web sites, desktop applications, and even files are types of information sources that can be related to particular events. Each type of information source requires a specific extractor, capable of retrieving information from the appropriate protocols and API. Source-specific rules define how the extractor should interact and extract data from a particular site. Desktop indexing applications, such as Google Desktop, may facilitate this type of extraction, providing a single point of integration for multiple sources.

Information extraction must be flexible to support different types of relationships, dealing with dynamic content from diverse sources from the Web and desktop applications. Relationships can range from recent emails from an event attendee to news and blog entries that contain keywords in common with the event to traffic near a particular location. A flexible mechanism is needed to allow the user to define the relationships of interest for types of events, along with useful default rules. Search technologies clearly can be applied here as a mechanism of extracting and identifying related information. Of course, not all information need be extracted, often a reference or hyperlink to the information will suffice.

2.3 Impact Assessment

A number of factors must be considered when determining the severity and urgency of information. The purpose of assessment is to determine the degree to which an item of information will impact an event. The most basic assessment method is determining severity from the source of information. Information from particular sources is assigned a fixed severity, such as all traffic information being given. The type of information and The type and source of information provide a first approximation of the degree of impact,

A second measure of impact is the content of the information, whether it contains particular keywords or phrases. A third measure of impact is how it relates to the event, as an impact that relates to an event based on Some sources may always provide important information, such as a traffic site, whereas other sites

Most importantly is the fact that degree of impact is subjective, varying by user and therefore requiring that any impact assessment must take into account the user's personal preferences. Several of the item of Given that each type of information generally impacts an event along one factor. Travel related information, traffic and weather, will generally travel, whether driving and air travel, can be assesFor instance, inclement weather may significantly impact a long drive, but have a minimal impact on a short drive. A straightforward approach to determining severity is to apply rules, specific to each information source, to assign severity to information. These rules should take into account the information and the event itself, such as "if weather is heavy snow and traveling by car, then severity is critical." The focus of impact assessment is the accuracy of the severity and urgency for a given information item. Impact assessment relies on an augmented calendar constructed through context-aware retrieval, as illustrated in Figure 1.

Severity is dependent on a number of factors concerning the information and the event in the user plan. A travel event, for instance, is impacted by information that changes the scheduling or duration of the event. Bad weather, traffic, road construction, and the like can all affect the duration of driving between two locations. In addition, impacts can have a compounding effect. Delayed travel can impact a subsequent event, requiring that the user either leave earlier or reschedule.

The concept of impact is not just a question of where, when and how long an event takes place. Impact must take into consideration whether the information will affect how the user performs the event. Press releases, research papers, blog entries, and news articles can all allow the user to perform an event better. Newly available information can be directly related to some event the user is performing, whether it is writing a paper, researching a topic, attending a meeting or making social plans.

After an assessment of severity, the urgency of the information must be determined, to evaluate when and how the user should be notified. To make this assessment, the severity, proximity in time, the user's current context, and the type of information may be taken into account. Immediate notification should be reserved for only the most urgent information, that which the user must be made aware of immediately, regardless of the user's current context. Delayed notification should be used to interrupt the user at the next convenient point in time. Severity and urgency must also be tunable by a user, allowing the user to opt for fewer or more notifications, or suppress notifications of certain types of information.

2.4 Information Delivery

Improving awareness requires timely delivery of information to the user. One definition of information awareness is "the amount of time between the availability of new information and the time at which the user acknowledges or consumes this information"[5]. Yet, timely delivery of information must be balanced with the need to minimize the significant negative impact of interruptions [2]. Frequent interruptions can lead to users becoming conditioned to ignore or suppress the interruptions, leading to a reduction in information awareness, and should be avoided. Figure 1 shows several potential interfaces, from a mobile phone to a web browser to a calendar application.

Knowledge of future context is required to balance the need for timely notifications with the goal of minimal disruptions. Interruptions can be very disruptive, requiring the user to focus their attention on the new information and away from the task at hand. Urgent information pertaining to imminent events, such as a meeting cancellation, or representing significant issues, such as a hurricane warning, warrants an immediate interruption whereas less urgent information may be delayed to reduce disruption. While knowledge of current context alone allows the delay of interruptions until the user is not busy, current context does not provide a basis for inferring when the user will be free and thus is insufficient for determining if delaying an interruption for a short time will be helpful. Selecting the appropriate time to deliver information to the user requires balancing what the user is currently doing and what the user will be doing, as well as the urgency of the information itself. Notifications delivered by a context-aware retrieval system may therefore employ disruption avoidance techniques, allowing high disruption for only severe information that impacts an imminent event.

Much information will not require the user's immediate attention, as reflected by a low urgency. This information should be made accessible to users via an interface that allows browsing and inspection of events with related information. As a calendar is the primary interface for managing events, integrating related information into the calendar application may provide the best user experience. On the other hand, as calendars are not ubiquitous and synchronization of calendars with mobile devices is today limited, especially on non-smartphones, a web-based interface that provides the user plan along with related information would also be an effective mechanism. Information that should be browsable in the context of future events not only includes the potentially impacting information, but also reference information. Examples of reference information include directions to the event, phone numbers of attendees, personal websites, and hotel reservation information.

3 The IMPACT System

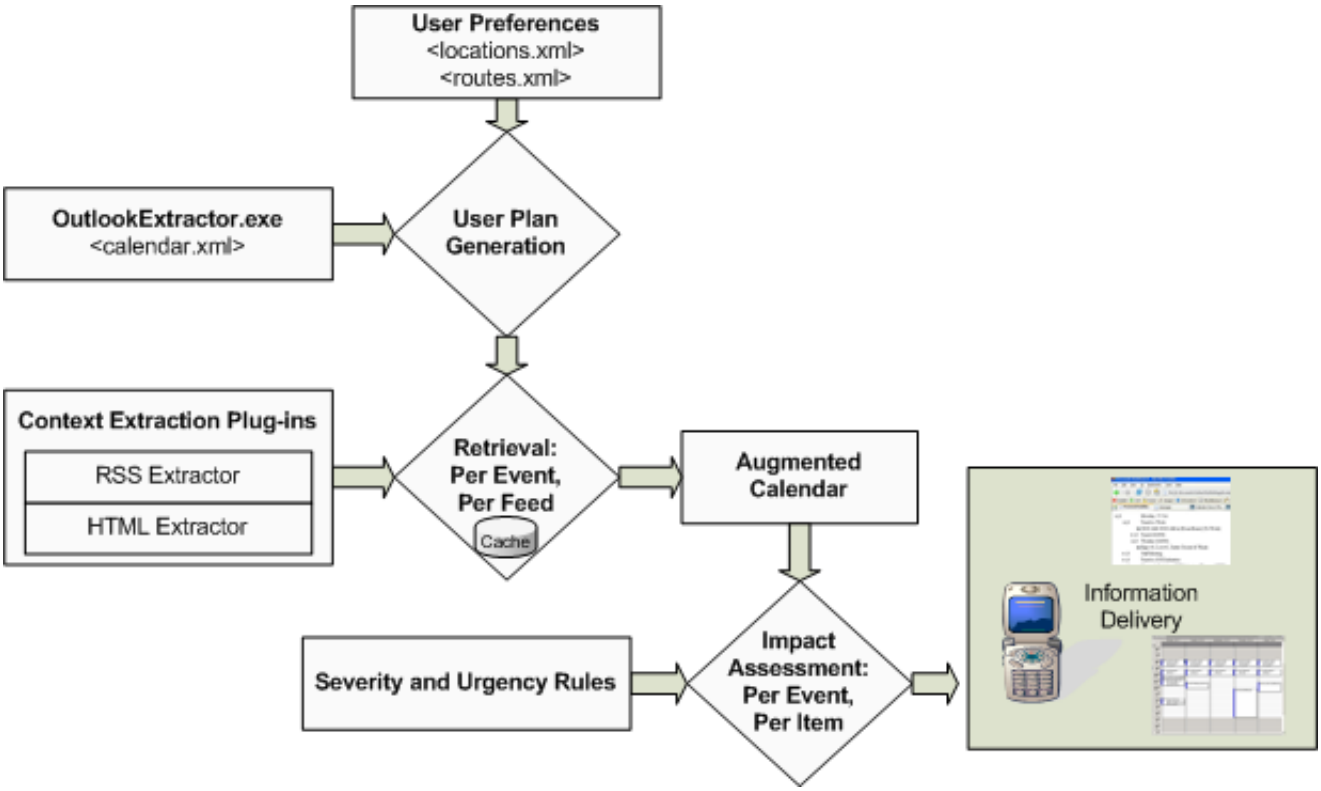

Figure 2: Overview of the IMPACT System.

To explore the concepts discussed in the previous section, the Integrated Management of a Personal Augmented Calendar Tool (IMPACT) was built. IMPACT is a web-based Java application that demonstrates the feasibility and highlights the challenges of context-aware retrieval using future context.

Following from the conceptual view in Figure 1, Figure 2 illustrates the high level architecture of the IMPACT system. IMPACT is made up of four high level components:

User Plan Generation, Context-Aware Retrieval, Impact Assessment, and Information Delivery. User plan generation employs user calendar information to determine the user plan. Context-aware retrieval relates information from information sources to the user plan. Impact assessment determines the severity and urgency of each related information item. Finally, information delivery presents information to the user. Each of these are discussed in the remainder of this section.

IMPACT is able to relate email, traffic, weather, and news articles to particular events in the user calendar. Relationships are based on the timeframe, location and keywords for a given event. Each related information item is assessed for severity of impact based on keywords contained in the information, such as "Flood Watch" in a weather forecast or "Accident" in a traffic report. Urgency is then determined as an offset from the severity, either increasing or decreasing the severity, through a combination of the user's current context and the time before the event starts. Items relating to imminent events have increased urgency, and items in the further future have decreased urgency. Finally, information is presented to the user through a web-based interface and through email notifications.

3.1 User Plan Generation

The basis for the user plan in the IMPACT system is an electronic calendar application, Microsoft Outlook. As shown in Figure 2, user preferences and calendar information are combined to form the predicted context. For this purpose, a Microsoft Outlook calendar extractor has been written to export calendar events in a structured (XML) representation, which serves as input to the user plan generation engine. A typical calendar event contains the following information:

- subject,
- start date/time,
- end date/time,
- unique id,
- location, and
- attendees.

The user plan generation engine periodically receives the calendar data, converts them to an internal representation and expands events based on user-defined preferences. These preferences include location details, such as the address for a particular location, and route information for travel between locations. To facilitate the relation of preferences to calendar data, some conventions are used to simplify parsing. For instance, locations are specified within parenthesis, such as "$WPICampus$", and flight numbers are prefixed with "Flight:", such as "$Flight : AA75$". These conventions are merely for prototype purposes, and are intended to be generalized to eliminate the additional calendar management effort.

Currently, the user plan generation engine does not predict new events, however it is envisioned that it will analyze the user's calendar, leverage history, and to do items, to

predict unlisted events and assess the likelihood of the user participating in an event. In addition, pre-processing of the calendar is expected to be required to compensate for implied travel and other inconsistencies.

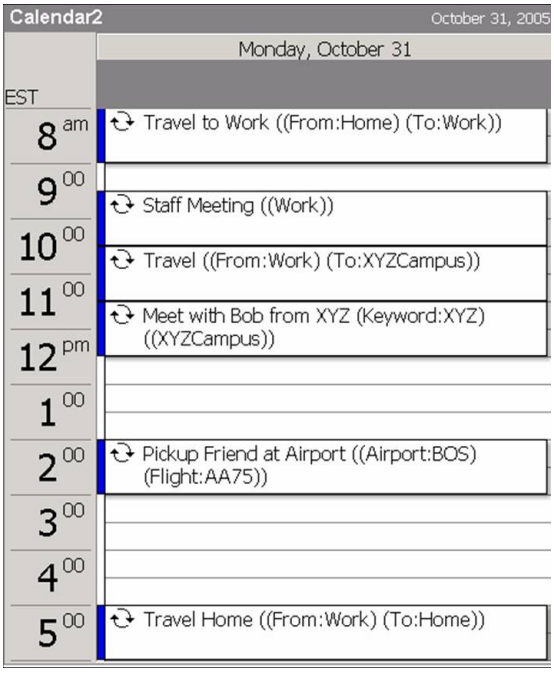

Figure 3: Sample calendar.

As an example, Figure 3 shows a user calendar, beginning with a morning commute from home to work. Monday begins with a staff meeting, and then traveling to Company XYZ's campus for a meeting with Bob at 11:30. Later in the day, the user will pick up a friend at the airport on Flight AA75. After work, the user commutes home.

The output of user plan generation is the user plan, comprised of calendar events with additional details filled in from user preferences. This user plan is provided as input to the context-aware retrieval engine.

Figure 4 illustrates the resulting user plan, based on a portion of Figure 3. In this user plan, the user commutes to work daily at 8:30am, and as a morning staff meeting at 9:30. Two attendees are invited, Tom and Mary. The user also has a meeting at 11:30 with Bob from XYZ.

3.2 Context-Aware Retrieval

With a generated user plan, information is then extracted and related to each event by the retrieval engine. Extraction of information is performed by protocol-specific plug-ins, as illustrated in Figure 2. Each extractor applies protocol specific logic to select, extract and filter content, producing information items that may be related to events. Each information item contains a textual summary, URL, date range, and location. Items that do not have an explicit date range provided by the source are, depending on the source, given either an indefinite timeframe with no end date or applied to the current day.

0830-0850	Travel to Work
Location	Home to Work
Distance	20 miles
Route	Main St to Rt 90 to James St
0930-1030	Staff Meeting
Location	Work
Attendees	Tom (Organizer), Mary (Declined)
1130-1230	Meet with Bob from XYZ
Keyword	XYZ
Location	XYZCampus
Attendees	Bob (Accepted)

Figure 4: User Plan Example

Three extractor plug-ins have been written for the IMPACT system: HTML, RSS and Microsoft Outlook Email. The HTML extractor is oriented towards extracting specific blocks of text from documents, and employs a DOM-based technique whereby a well-known anchor point is searched for and a DOM offset from the anchor is applied to find the target content. HTML sources that have been demonstrated include flight status, airport status, stock price, and news sites. As RSS's structure is well-known, the RSS extractor will normally extract information items directly from RSS items. Some sites embed multiple information items within a single RSS item, such as hourly forecasts within a single daily forecast RSS item, the RSS extractor must also support extracting elements from within items. RSS sources that have been demonstrated include weather and traffic RSS feeds. The email extractor retrieves unread emails from attendees, and is currently extracted from Outlook along with the calendar information. For each event, the user's email inbox is searched for unread messages from event attendees. Keyword and subject filtering are planned to be added in the future.

Content extraction is not the focus of this work, therefore we do not describe the details of the extractor plug-ins. Any method of extraction that can provide content along with associated context can be used in the IMPACT system, including work in [15, 19].

In RSS and HTML sources, selection is performed through parameters in a URL query string. For example, to extract traffic information from Yahoo, the following components are needed to construct the URL:

- Base URL: `http://maps.yahoo.com/traffic.rss`
- Parameter: csz
- Value: event.location.zipcode

As a last step, each extracted information item is checked for relevance to a given appointment. Even though the URL for the information item was constructed based on context from the appointment, often multiple information items are returned that partially map to the given appointment. For instance, weather forecasts are extracted for each zip code, yet

multiple forecasts are provided containing different time ranges, of which at most one will apply to a given event. To minimize the burden on web sites due to repeated queries, information is cached locally for each URL for a configurable time period. This process is repeated periodically to check for new information, and eliminate stale information.

0830-0850	Travel to Work
Location	Home to Work
Distance	20 miles (directions)
Route	Main St to Rt 90 to James St
Traffic	Main St: Construction
Traffic	Rt 90: Accident, One Lane Open
Weather	Partly Cloudy & Windy, Hi 60^o, Low 41^o
0930-1030	Staff Meeting
Location	Work
Attendees	Tom (Organizer), Mary (Declined)
Emails	1 from Tom
1130-1230	Meet with Bob from XYZ
Location	Work to XYZCampus
Keyword	XYZ
Attendees	Bob (Accepted)
Emails	2 from Bob
News	"XYZ Releases v2.0 of XYZSoft"

Figure 5: Augmented Calendar

Figure 5 illustrates an augmented calendar, the result of context-aware retrieval. The morning travel to work event has traffic and weather associated with it. The morning staff meeting is related with a single email from the organizer. The 1130 meeting with Bob was related to a news headline based on the XYZ keyword in the event. In addition, two unread emails from the attendee, Bob, were found.

3.3 Impact Assessment

Impact assessment applies user specified severity and urgency rules to the augmented calendar, as shown in Figure 2. Each information item in the augmented calendar is evaluated for severity, and then urgency. The IMPACT system uses a simple keyword-based approach for assessing severity of impact with different keywords for each source. A set of pattern-based rules compare attributes of the event, the information, and possibly externally parameters for each information/event pair. This level of impact is a rating, from most impacting to least: critical, medium, low, none. For example, if the weather forecast contains "heavy snow" or "flood watch", then the forecast is of critical severity. While severity rules are based on the information itself, the system could be extended to consider additional details from the context, such as light snow in Florida being more severe than heavy snow in Alaska.

While severity only takes into account the information item, and possibly the appointment context, urgency considers the user's overall context. The urgency of an event begins as being equal to its severity, and is then increased or decreased based on two factors: time-to-event and the user's current context. The time-to-event refers to the amount of time until the appointment is scheduled to begin. The user's current context is defined as being either "idle", "busy", or "transit". The closer the event, the more urgent the information is, whereas the further away in time the event is the less urgent. Likewise, when the user is busy, the urgency of information is reduced significantly to compensate for the disruptiveness of interruptions [17]. Note that some information sources are considered as being never urgent, such as those that are for reference purposes only.

0830-0850	Travel to Work
Traffic	(Medium) Main St: Construction ...
Traffic	(Critical) Rt 90: Accident ...
Weather	(Minor) Partly Cloudy & Windy ...
0930-1030	Staff Meeting
Emails	(Medium) 1 from Tom
1130-1230	Meet with Bob from XYZ
Emails	(Medium) 2 from Bob
News	(Minor) "XYZ Releases v2.0 of XYZSoft"

Figure 6: Augmented User Plan with Severity

Figure 6 expands on the example from Figure 5. The morning commute to work has a critical item, an accident on Rt 90, whereas the construction on Main St is only of medium severity. Both the accident and construction are assessed based on keywords, for example the occurrence of the word "accident" in traffic will always be marked critical. In the meeting with Bob from XYZ, unread messages from an attendee are rated as medium severity.

3.4 Information Delivery

The IMPACT system provides a web-based dashboard for browsing of the augmented calendar and delivers email notifications of urgent information. Figure 7 is an example of the dashboard, following from Figure 6.

The dashboard provides an integrated view of the user's calendar details along with the related information items for each event. Items above an urgency threshold are marked with an exclamation point icon to highlight the items requiring the user's attention. An alternate approach would be to augment the calendar application itself to provide an integrated view for users. For portability and rapid prototyping purposes, a web application was chosen. Most mobile devices provide web browsers, allowing demonstration of the IMPACT system from a wide range of devices. The web-based approach comes at the cost of providing the user with a secondary calendar interface, adding the risk of inconsistencies with the primary calendar application.

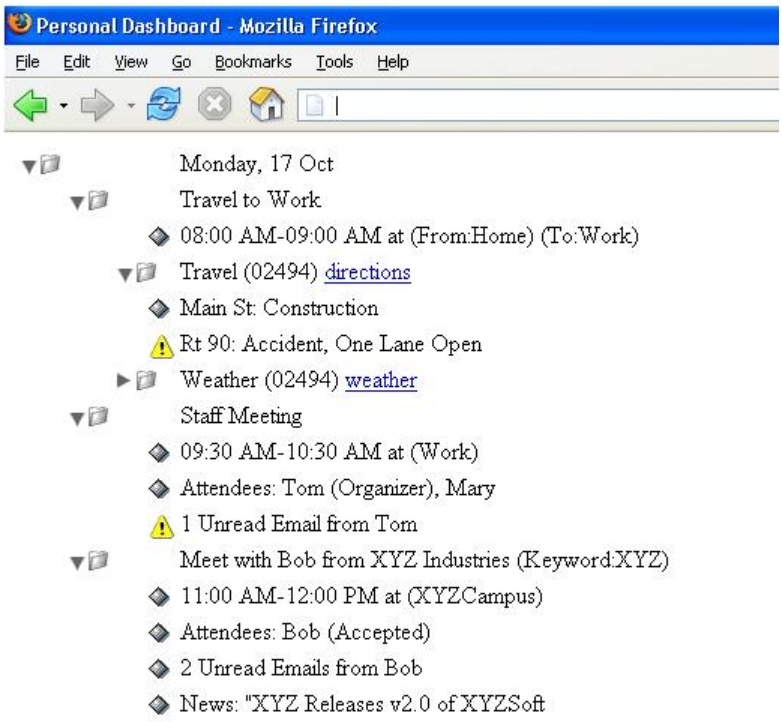

Figure 7: IMPACT Dashboard.

Email notifications are used to notify the user of urgent information, providing the details of the information along with the related event. When the urgency for an information item changes, or a new information item is related to an appointment, it is evaluated to determine whether it should trigger a notification. This determination is based on the urgency of the information, with a user-definable threshold to establish the minimum level of urgency that requires a notification. The default threshold is for critical events to be sent immediately. Both the dashboard and email notifications are accessible from mobile phones, as many mobile phone carriers provide mobile web browsers and email to text message gateways.

Three parameters are used to control the frequency and content of notifications: immediate urgency threshold, secondary urgency threshold, and maximum items per notification. The immediate urgency threshold denotes the urgency at which items will be sent immediately as a notification. The secondary urgency threshold is used to maximize the user's attention by sending lesser urgent information along with information that meets the immediate urgency threshold. The secondary urgency threshold is based on the expectation that some additional information in an interruption will not result in significantly higher disruption. Finally, the maximum items per notification limits the number of secondary items, to prevent a large number of items being sent in a notification.

4 Evaluation

While we have not formally evaluated the IMPACT system, we have clear ideas on how this evaluation should be done. This section describes methods for evaluating each of the components of the IMPACT system shown in Figure 2. As previously stated, the objective of the IMPACT system is to achieve the highest recall and precision with minimal disruption. The IMPACT system provides a focal point for information that is likely to be accessed, organizing information according to the user plan and interrupting the user with critical information only when necessary.

The context awareness of the IMPACT system depends on accuracy of the user plan, thus depending on the accuracy of the user calendar. If the user calendar is a poor predictor of user behavior, then it is likely that the resulting information will be poorly matched to the user's interests. The accuracy of the calendar in predicting user behavior may be measured through user tracking, user feedback, or a combination of the two. To get user feedback, the IMPACT system may periodically ask a user to annotate previous events with whether they were attended, skipped, changed, or for reference-only. Tracking of user current context, including location and presence, allows analysis of whether the user plan accurate predicts user behavior.

High recall in context-aware retrieval should result in a reduced need to periodically check information sources, and fewer occasions where the user is not aware of an information item that is related to an event. Measuring the change in email and web browsing behavior as a result of the IMPACT system should indicate a reduction in periodic browsing. To assess whether the IMPACT system is missing information, the information delivery mechanisms should allow the user to mark particular events as "missing information", along with a URL for where the information can be found.

Evaluating the precision provided through impact assessment must also be performed through user feedback. Within the browser interface, the user should be able to adjust the severity and urgency of events. Adjustments would be recorded for further analysis to determine the overall precision of the assessments, and whether inaccuracies are the result of particular rules or if the overall approach of assessing severity must be changed. A more general measure is whether critical events do in fact result in a change in the user behavior, and whether there is any correlation between particular events and particular changes in user behavior. For example, do users leave early when informed of a critical traffic issue or poor weather?

Finally, the disruptiveness of notifications would be assessed by allowing user feedback upon interruption. A user should be able to flag unnecessarily disruptive notifications by replying to the email. In addition, users should be able to record the disruptiveness of various notifications. These measures would provide an indication of whether the urgency assessment and subsequent notification rules provide an effective means for reducing the disruptiveness of interruptions.

5 Related Work

There are a number of research areas related to this work, ranging from context-aware computing to mobile computing to personalization research. The use of history and other mechanisms to predict future context based on history and predictions [26] may improve the accuracy of the user plan, providing a potentially more complete user plan than one based solely on user calendars. [22] explores using calendars as a source of context for mobile devices. Information retrieval research has explored the role of context awareness, and how to account for issues of changing user context in retrieving and filtering information [7].

The use of future context for interruption management is explored in [20], which presents "methods for inferring the cost of interrupting users based on multiple streams of events ... and data drawn from online calendars." [17] observes that interruptions are "better received" during physical transitions, such as standing up. The use of physiological cues, such as heart rate, to reduce the disruptiveness of interruptions is explored in [10]. When the user is at a desktop or laptop, [5] suggests the use of popups and message boxes for gaining the users attention with minimal intrusion.

A significant challenge in information retrieval is extracting meaningful information from information sources. While the dominant method of distributing content, HTML is a presentation-oriented structure, making automated extraction difficult. Automated HTML extraction is explored in [19, 13]. RSS is a commonly used XML-based format for summarizing web content which facilitates information extraction, although without providing the meaning of the contained information [24].

The Semantic Web initiative takes a different approach. Instead of inferring meaning from pages, information should be provided by web sites already organized and structured according to its meaning. By associating not only structure, but also meaning, to web content, the Semantic Web provides the foundation for richer applications to be developed around the web [25]. When, and if, RDF content becomes widely available, the challenge of information extraction will be greatly simplified [16], but in the meantime any approach must provide for extraction of information from a wide range of sources, including HTML, RSS and web services. Other sources of information appropriate for personal information retrieval also include email, files, and databases.

Unplanned activities can change the user plan unexpectedly and thus render it irrelevant or inaccurate. Research has found that unplanned activities make up a significant amount of our daily interactions [21]. These activities present a difficulty as there is no basis for relevant information to provide to the user nor does the system know when the user can be interrupted. Methods can be applied that minimize potential disruptions, such as letting a caller view what periods the callee is busy and free [9].

Mobile devices, including phones and personal digital assistants, now provide nearly universal access to a wide range of content. They are an important tool used to communicate and access information. Yet, the availability of content for these devices is quite limited and difficult to use [12]. To overcome mobile browsing issues, [8] presents a method that summarizes web content for mobile devices. [6] comments on the general problem of mobile web access, pointing out that: "A literal translation from the Web to wireless is inadequate. Merely squeezing data onto small screens detracts from the user experience on mobile devices. By adaptively learning users' preferences, all users can have easy access to a vast amount of

information at any time."

Context-aware computing research has focused on how knowledge of the user's situation, such as location and presence, can be leveraged to enhance the computing experience [11]. Physical location is a broadly useful source of context [23], which can be used to enhance a user's web experience through location-aware applications. For example, [27] provides a service that delivers notifications based on the location of the user. An event planner that leverages contextual information to make more informed decisions is presented in [28]. Context-aware research highlights the value of integrating diverse sources of context to enable new types of applications, applications that can adjust to the user's changing needs and interests.

Work in personalization provides methods to personalize web content by learning user's interests and adapting web sites to fit the user [3]. A related argument is that that a one size fits all web site model, which can be adapted dynamically for desktop and mobile users, will not address the different needs of mobile users. Instead, [4] presents a web site personalizer that "automatically adapts and personalizes a web site." [4] also note that users exhibit different goals and behaviors, from goal-oriented searching to browsing/surfing.

6 Summary and Future Work

Improving information awareness is an important goal and challenge in today's increasingly connected world. As information sources continue to expand, and users become universally connected to information, attention becomes a critical and limited resource. Maximizing user's attention, and reducing the effort required to monitor information, will continue to be a challenge in the years to come.

This work describes an approach for improving awareness through personal information retrieval, leveraging the user plan to determine the relevancy and importance of information. Improving information awareness requires the delivery of all information that will impact the users activities, requiring both high recall of information and precision in assessment of impact. In addition, the information delivery mechanisms must balance the need for immediate awareness with that of minimizing disruptions.

The IMPACT system demonstrates these concepts, providing a framework for aggregating information that is related to events in the user's predicted context. The user's future context, the user plan, is determined from calendar information, with events in the user plan related to information through rule-guided extraction performed by HTML and RSS extractors. Information is delivered to the user through emails and a web portal, allowing both notifications and browsing of information.

Several capabilities are planned to be added to the system. Providing control to the user through a user interface over retrieval and impact assessment will minimize the technical competence required by the user's being studied. Allowing users to add new sources, modify relationship determinations, and change severity calculations, will allow the system to be customized by the end-users. In addition, minimizing the burden of formatting calendar events in particular ways, such as "(From:location)", will also minimize the effort required of the user to maintain the calendar properly.

Other related areas for future work, but outside the scope of this work, include the role

of Semantic Web technologies, including RDF query languages, and whether they provide improvements in the flexibility and expressiveness of relationships in the IMPACT system. The relationship and role of other individual's future context for a given user should also be investigated.

References

[1] ForecastFox. http://forecastfox.mozdev.org/.

[2] P. D. Adamczyk and B. P. Bailey. If not now, when?: the effects of interruption at different moments within task execution. In *CHI '04: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 271–278, New York, NY, USA, 2004. ACM Press.

[3] M. Albanese, A. Picariello, C. Sansone, and L. Sansone. Web personalization based on static information and dynamic user behavior. In *WIDM '04: Proceedings of the 6th annual ACM international workshop on Web information and data management*, pages 80–87, New York, NY, USA, 2004. ACM Press.

[4] C. R. Anderson, P. Domingos, and D. S. Weld. Personalizing web sites for mobile users. In *WWW '01: Proceedings of the 10th international conference on World Wide Web*, pages 565–575, New York, NY, USA, 2001. ACM Press.

[5] B. P. Bailey, J. A. Konstan, and J. V. Carlis. Adjusting windows: Balancing information awareness with intrusion. In *Proceedings of the 6th Conference on Human Factors and the Web*, 2000.

[6] D. Billsus, C. A. Brunk, C. Evans, B. Gladish, and M. Pazzani. Adaptive interfaces for ubiquitous web access. *Commun. ACM*, 45(5):34–38, 2002.

[7] P. J. Brown and G. J. F. Jones. Context-aware retrieval: Exploring a new environment for information retrieval and information filtering. *Personal Ubiquitous Comput.*, 5(4):253–263, 2001.

[8] O. Buyukkokten, H. Garcia-Molina, and A. Paepcke. Seeing the whole in parts: text summarization for web browsing on handheld devices. In *WWW '01: Proceedings of the 10th international conference on World Wide Web*, pages 652–662, New York, NY, USA, 2001. ACM Press.

[9] J. Cadiz, A. Narin, G. Jancke, A. Gupta, and M. Boyle. Exploring pc-telephone convergence with the enhanced telephony prototype. In *CHI '04: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 215–222, New York, NY, USA, 2004. ACM Press.

[10] D. Chen and R. Vertegaal. Using mental load for managing interruptions in physiologically attentive user interfaces. In *CHI '04: CHI '04 extended abstracts on Human factors in computing systems*, pages 1513–1516, New York, NY, USA, 2004. ACM Press.

[11] G. Chen and D. Kotz. A survey of context-aware mobile computing research. Technical Report TR2000-381, Dept. of Computer Science, Dartmouth College, November 2000.

[12] L. Chittaro and P. D. Cin. Evaluating interface design choices on wap phones: Navigation and selection. *Personal Ubiquitous Comput.*, 6(4):237–244, 2002.

[13] O. Etzioni, M. Cafarella, D. Downey, S. Kok, A.-M. Popescu, T. Shaked, S. Soderland, D. S. Weld, and A. Yates. Web-scale information extraction in knowitall: (preliminary results). In *WWW '04: Proceedings of the 13th international conference on World Wide Web*, pages 100–110, New York, NY, USA, 2004. ACM Press.

[14] H. Green. All the news you choose - on one page: Rss, which delivers customer-tailored bulletins to users, may shake up e-media. BusinessWeek, oct 2004.

[15] S. Gupta, G. Kaiser, D. Neistadt, and P. Grimm. Dom-based content extraction of html documents. In *WWW '03: Proceedings of the 12th international conference on World Wide Web*, pages 207–214, New York, NY, USA, 2003. ACM Press.

[16] J. Hendler, T. Berners-Lee, and E. Miller. Integrating applications on the semantic web. In *Journal of the Institute of Electrical Engineers of Japan*, volume 122, pages 676–680, October 2002.

[17] J. Ho and S. S. Intille. Using context-aware computing to reduce the perceived burden of interruptions from mobile devices. In *CHI '05: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 909–918, New York, NY, USA, 2005. ACM Press.

[18] J. Ho and R. Tang. Towards an optimal resolution to information overload: an infomediary approach. In *GROUP '01: Proceedings of the 2001 International ACM SIGGROUP Conference on Supporting Group Work*, pages 91–96, New York, NY, USA, 2001. ACM Press.

[19] A. Hogue and D. Karger. Thresher: automating the unwrapping of semantic content from the world wide web. In *WWW '05: Proceedings of the 14th international conference on World Wide Web*, pages 86–95, New York, NY, USA, 2005. ACM Press.

[20] E. Horvitz and J. Apacible. Learning and reasoning about interruption. In *ICMI '03: Proceedings of the 5th international conference on Multimodal interfaces*, pages 20–27, New York, NY, USA, 2003. ACM Press.

[21] E. A. Isaacs, J. C. Tang, and T. Morris. Piazza: a desktop environment supporting impromptu and planned interactions. In *CSCW '96: Proceedings of the 1996 ACM conference on Computer supported cooperative work*, pages 315–324, New York, NY, USA, 1996. ACM Press.

[22] A. Khalil and K. Connelly. Improving cell phone awareness by using calendar information. In *INTERACT2005: Tenth IFIP TC13 International Conference on Human-Computer Interaction*, 2005.

[23] T. Kindberg, J. Barton, J. Morgan, G. Becker, D. Caswell, P. Debaty, G. Gopal, M. Frid, V. Krishnan, H. Morris, J. Schettino, B. Serra, and M. Spasojevic. People, places, things: web presence for the real world. *Mob. Netw. Appl.*, 7(5):365–376, 2002.

[24] R. M. Lerner. At the forge: syndication with rss. *Linux Journal*, 2004(126):8, 2004.

[25] C. C. Marshall and F. M. Shipman. Which semantic web? In *HYPERTEXT '03: Proceedings of the fourteenth ACM conference on Hypertext and hypermedia*, pages 57–66, New York, NY, USA, 2003. ACM Press.

[26] R. Mayrhofer. Context prediction based on context histories: Expected benefits, issues and current state-of-the-art. In T. Prante, B. Meyers, G. Fitzpatrick, and L. D. Harvel, editors, *Proceedings of the 1st International Workshop on Exploiting Context Histories in Smart Environments (ECHISE2005)*, May 2005. part of the Third International Conference on Pervasive Computing (PERVASIVE 2005).

[27] J. P. Munson and V. K. Gupta. Location-based notification as a general-purpose service. In *WMC '02: Proceedings of the 2nd international workshop on Mobile commerce*, pages 40–44, New York, NY, USA, 2002. ACM Press.

[28] Z. Pousman, G. Iachello, R. Fithian, J. Moghazy, and J. Stasko. Design iterations for a location-aware event planner. *Personal Ubiquitous Comput.*, 8(2):117–125, 2004.

[29] V. Raghavan, P. Bollmann, and G. S. Jung. A critical investigation of recall and precision as measures of retrieval system performance. *ACM Trans. Inf. Syst.*, 7(3):205–229, 1989.