Delivering Real Reality to the World Wide Web via Telerobotics

Eric Paulos[*]
paulos@cs.berkeley.edu

John Canny[*]
jfc@cs.berkeley.edu

Department of Electrical Engineering and Computer Science
University of California, Berkeley

Abstract

Robots provide us with a means to move around in, visualize, and interact with a remote physical world. We have exploited these physical properties coupled with the growing diversity of users on the World Wide Web (WWW) [1] to create a WWW based telerobotic remote environment browser. This browser, called Mechanical Gaze, allows multiple remote WWW users to control a robot arm with an attached camera to explore a real remote environment. The environment varies but is typically composed of collections of physical museum exhibits which WWW users can view at various positions, orientations, and levels of resolution.

http://vive.cs.berkeley.edu/capek

1 Introduction

We have designed this teleoperated WWW server in order to allow users throughout the world to visit actual remote spaces and exhibits. It also serves as a useful scientific tool by promoting discussion about the physical specimens in the browser such as insects, live reptiles, rare museum collections, and recently discovered artifacts.

The use of an on-line controlled camera eliminates many of the resolution and depth perception problems of libraries of digitized images. The user has complete control over the viewpoint, and can experience the exhibit in its state at a particular moment in time, under the same conditions and lighting as a viewer who is in the actual space.

In addition, each exhibit has a hypertext page with links to texts describing the object, other web pages relevant to it, and to comments left by other users. These pages can be accessed by navigating the camera in physical space, and centering on a particular object. The pages can be thought of as mark-ups of 3D objects in the spirit of VRML [2], but where the objects are actual physical entities in a remote space rather than simply models.

Exhibits can be added or removed in a matter of a few minutes, allowing for an extremely dynamic array of objects to be viewed over the course of only a few months. Users are encouraged not only to check back for upcoming exhibits, but to participate themselves. Users can leave commentary about an item on exhibit, creating dialogue about the piece, as well as feedback to the owner, artist, or curator of the object. Institutions, museums, curators, scientists, artists, and individual users are all invited to exhibit objects.

2 Goals and Motivation

Initially, we were driven to develop a useful application for interactive telerobotics. We were inspired by the diversity and growth of the WWW as the medium for such an inexpensive, publicly accessible tool for remote environment browsing. The restrictions imposed by the Hyper Text Markup Language (HTML) made it difficult to design an intuitive user interface to a complex robotic system. Certainly, we could have chosen to construct custom navigation software for users to download. While this would allow us more freedom in the design of the overall system, it would severely restrict the accessibility of the system. Since we consider the quantity and diversity of users on the WWW as one of its most powerful aspects, we chose to constrain the development of our system within the accessibility of WWW users.

2.1 Goals

Before designing the system we set forth our goals for the project. Our primary goal was to provide a universal remote environment browsing tool that is useful for the arts, sciences, and in the development of education and distant learning. To meet this goal we agreed upon several elements that we felt were essential to any WWW based telerobotic system.

First, we wanted to insure universal unrestricted access to the system. This would allow access to artifacts and objects by a wider audience than previously available. Current access restrictions are usually the result of geographic, political, or monetary constraints preventing the individual from traveling to the object. Likewise, owners and curators of exhibits do not always have the resources or the desire to tour the objects throughout the world. We wanted to develop a tool that would attempt to solve many of these problems by

bringing the people together with the objects at a minimum cost.

Rather than a fixed, static display, the browser must allow users true three-dimensional navigation around objects at varying positions, orientations, and levels of resolution. As David Gelernter suggests in his book *Mirror Worlds* [3], such systems that gaze into remote spaces should show each visitor exactly what they want to see. This requires the system to provide millions of different views from millions of different focuses on the same object. Certainly visitors will desire to zoom in, pan around, and roam through the world as they choose. More importantly, they should be permitted to explore this space at whatever pace and level of detail they desire. Users should also be free to swivel and rotate the image, to get a better look at regions that might be obscured in the initial perspective.

The telerobotics browser should also provide to the exhibit owners, curators, and caretakers a forum to receive feedback and commentary about their exhibit. This same forum should also allow scientists to discuss details concerning classification of specimens such as insects or the origins of an recently discovered artifact. Essentially, some method for leaving comments and creating dialogue should be provided.

Finally, the system should allow exhibits to be added and removed with a minimum of effort, thus providing the possibility of exhibiting a wide variety of objects over the course of a few months. In addition, recently discovered/developed scientific objects should be able to be added for universal browsing within the order of a few minutes.

2.2 Why Use Live Images?

A common objection to our approach is why we simply do not use pre-stored digitized images for browsing objects and spaces. While we agree upon the importance of such pre-stored images, the remote environment browser offers several distinct advantages over conventional image database solutions.

For example, the standard approach to providing remote access to museum collections' visual data is to digitize and pre-store images of all artifacts or specimens. This solution requires considerable expense and time commitment to complete the capture, storage and serving of digitized images. Our telerobotic approach allows remote scholars to interactively view museum artifacts and specimens on demand. This allows them to achieve much higher image resolution without the expensive digital storage requirements typically associated with large image databases that support distributed users and offer high resolution. Our interactive viewing solution also relieves museums of the need to store digital images of entire collections over a variety of resolutions.

Our approach allows immediate visual access to any/all collection materials from the beginning of a project. Traditional image capturing can take several years for large research collections, with millions of specimens that require special handling. The remote environment browser solution eliminates the waiting period that usually occurs during serial indexing and image capture. Museums that utilize a remote browsing model are able to provide remote access to any/all of their collection materials at a moment's notice, as opposed to access to a serially increasing number of objects over time. The ability to view specimens is more valuable if all specimens are available, the fewer specimens in a collection that are digitized, the less research value accrues to the resource as a whole.

By allowing researchers to choose their own view and magnification of the specimen or artifact, arguments over which specific view or number of views a museum should provide to remote users are eliminated. With a three dimensional object there will always be arguments surrounding what view to capture. Unless users can choose their own view of museum collections' materials, they will not be satisfied with using digital images for research. Even more importantly, some visually oriented research uses, such as taxonomy and morphology can not be supported in the digital environment without the provision of multiple views and magnifications. Useful statistics can be gathered by the browser as to which views are more popular among scientists and hence draw conclusions as to the relative importance of particular views and resolutions.

Certainly, dynamic exhibits such as live creatures, moving liquids, and mechanical systems must be viewed using live images. These live views are necessary to study the behavior of such systems.

Further discussions about the use of digital images in art and science, as well the implications of their use can be found in several sources [4; 5; 6; 7].

3 Previous and Related Work

The sensation of embodiment of an individual in a real life distant location has provided more than enough impetus for people to develop remote telepresence systems.

One of the earliest electrically controlled mechanical teleoperational systems was developed by Goertz [8] in 1954. Many subsequent systems were aimed at safely exploring hostile remote environments such as battlefields, nuclear reactors [9], deep oceans [10], mining [11], and outer space [12]. Additional applications for teleoperated surgery [13] and manufacturing [14] have been explored by several researchers [15; 16; 17].

Most of these system are quite complex, requiring special purpose dedicated hardware to control and interact with the

mechanism in the remote environment. As one of our goals states, we wanted to constrain development to a system that would be accessible to a wide audience without additional expensive or extraordinary hardware.

The spontaneous growth of the WWW over the past several years has resulted in a plethora of remote controlled mechanical devices which can be accessed via the WWW. Some of these early systems employed fixed cameras in remote spaces where users could observe dynamic behavior such as the consumption and brewing of coffee in a coffee pot or the activity of a favorite pet in its native habitat.

Systems evolved to allow users various levels of control via the WWW such as the LabCam [18] developed by Richard Wallace. His system allowed remote users to aim a pan/tilt camera using an intuitive imagemap interface.

Progression to intricate control of more degrees of freedom was realized by introducing robots to the WWW. Ken Goldberg *et al.* [19] developed a 3 DOF telerobotic system where users were able to explore a remote world with buried objects and, more interestingly, alter it by blowing bursts of compressed air into its sand filled world. Mark Cox [20] developed a system for allowing users to request images from a remotely controlled telescope. Another remote robotic system, developed by Ken Taylor [21], allowed WWW users to remotely manipulate blocks using a robot with an attached gripper. More recently, Ken Goldberg *et al.* have developed a telerobotic system called the TeleGarden [22] in which WWW users are able to observe, plant, and nurture life within a living remote garden. As of this writing, well over several hundred interesting mechanical devices are connected to the WWW with more spawning daily.

4 Overview

Our design choice for the user interface to the remote environment browser was to mimic much of the look and feel of a museum. We choose this approach, hoping that users would find it familiar to navigate, and thus more intuitive and inviting to use.

As a user enters Mechanical Gaze, they are presented with a chance to view some general information about the project, receive a brief introduction, obtain help in using the system, or enter the exhibition gallery.

Users who enter the exhibition gallery are presented with an up to date listing of the exhibits currently available for browsing. These are the exhibits that are physically within the workspace of the robot and can be explored. The idea behind the exhibition gallery is to give only a brief introduction to each of the available exhibits. This typically consists of providing the name of each exhibit, the dates it will be

available, the presenter(s), and perhaps a very brief description.

Users who wish to more closely examine an exhibit can simply select it from the listing. The user will then be presented with a more detailed description of the exhibit as well as a chance to either browse the exhibit using the robot or request to view the comments corresponding to that exhibit.

5 Hardware

The system has operated from both an Intelledex 605T robot with 6DOF and a 4DOF RobotWorld robot. The only noticeable external difference is that the roll and pitch operations are only available when the system is operating from the Intelledex robot. Otherwise, the back-end robotic hardware change is transparent to the user, giving hope to the prospect of such systems running on a variety of different robots in various environments.

Image capturing is performed using a camera and frame grabber hardware. Images are received from an RCA Pro843 8mm video camera mounted onto the last link of the robot. The auto-focus feature of the video camera allows users to view a variety of objects clearly, irregardless of the object's own height or the distance from which it is viewed. Typical exhibition spaces allow users to capture clear images anywhere from 3–30 cm from the surface of the object. Since we desired an easily reconfigurable exhibition space, a fixed focus camera would not be able to accommodate the wide variety of differently sized objects.

Image digitization occurs on either a VideoPix frame grabber card attached to a Sun IPC workstation or standard image capture hardware available on an SGI Indy. Eight bit 320x240 color images are captured in less than 50 ms. Further computation to convert the image into a compressed JPEG format for incorporation into HTML documents and save it to disk takes an additional 2–3 seconds. Overall, the time required to capture, convert, and save an image is on the order of 2–3 seconds.

The actual Hyper Text Transmission Protocol (HTTP) server containing the custom Common Gateway Interface (CGI) scripts and state information for individual users operates from an HP 715/60 workstation. This machine provides the front end interface to the system by receiving requests from WWW users, employing the services of the other hardware in the system, and delivering the results back to the user in an HTML format.

6 Robot Interface and Control

To interface the robot to the WWW, two separate pieces of code were written. The actual robot motion is performed by a daemon that accepts standardized requests via a socket
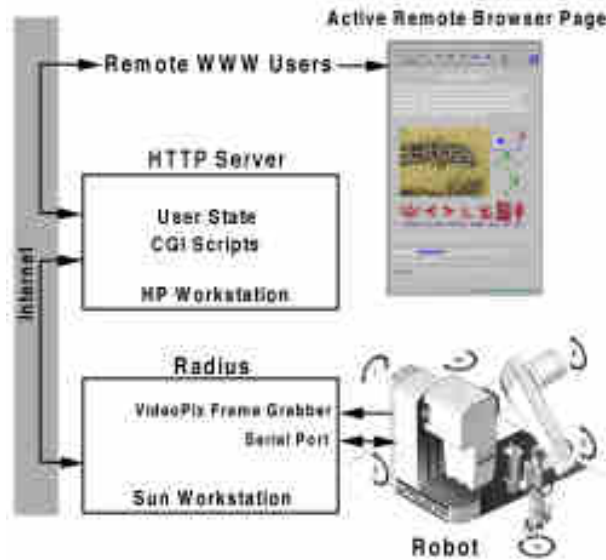
Figure 1: System Architecture with Intelledex Robot

connection and converts them into the native robot dependent commands. The other code interacts directly with the remote WWW user by handling administrative issues, resource contention, HTML page layout, and requests to the robot daemon when the robot must be moved.

6.1 Radius: The Robot Control Daemon

Radius, named after the main robot character in *R.U.R* by Karel Čapek [23], is the robot control daemon that provides a standardized interface to the various robots involved. By standardizing this interface, the rest of the system can be written ignoring any special kinematics or control systems for the particular end robot. Requests that involve control of the robot or camera hardware are handled by Radius. Radius listens for these requests on an established socket port.

When a socket connection is made, Radius first checks for authentication using a known encoding. This prevents unauthorized control of the robot hardware. This is particularly important as we move towards devices with the capacity of physical manifestations of energy in a remote environment. The damage resulting from an unauthorized access into such as system can cause not only irreparable damage to the robotic equipment and exhibits, but human injury as well. Therefore, measures to prevent at least the most naive attacks should be included in such systems.

Authorized connections to Radius include a 4 byte message. The message encodes the type of request and a mask. The request type can be a motion command or image capture command. This is followed by several bytes of data depending upon the request type and mask. Radius can also query

the robot to determine when all motions have stopped, hence allowing an image to be captured.

When an image grab request is received, Radius will employ the use of the available image capture hardware to capture an image, convert it to a 320x240 8 bit color JPEG image, assign it a unique identification number which is embedded into the image filename, and output it to a temporary space. The unique image number is passed back to the requesting process so that the correct corresponding image will be displayed in the HTML document delivered to the user.

Since our interface design is WWW based, requests are event driven. After a user has loaded an image, the robot is left idle until the user makes another request. Instead of allowing this exclusive access to the robot, leaving the robot idle while the user contemplates the next action, we service additional requests from other users. By multitasking, we provide increased access to the robot as well as a more efficient use of system resources. However, we must provide a method to guarantee that certain atomic operations are exclusive. For example, a request to move and grab an image, must be exclusive. This insures that no other motion occurs between the time we move the robot and capture the image. If we had failed to implement this, we would have no guarantee that the image delivered back to the user was actual taken from the location that they requested. The socket connection provides the mutual exclusion necessary to insure the correct functionality of Mechanical Gaze even when handling multiple requests. When a request is received by Radius, subsequent requests are queued until the first request has been handled. This insures that requests occur in order and have exclusive access to the robot and camera hardware for the duration of the request.

6.2 Navigation Page Construction

Requests to browse an exhibit are handled by a custom CGI script. Initially, the script is passed a unique identifying internal number corresponding to the exhibit to be browsed. The script reads in the current list of exhibits and extracts the relevant information for the exhibit of interest. One of these items is the physical location of the exhibit in the remote environment. Using this information, a socket connection is opened to Radius, the robot control daemon, and a request made to move the robot to the desired location and capture an image.

When the result of that request is received, the CGI script dynamically lays out the HTML page. First, it extracts information from the internal list of exhibits. This provides the name of the HTML file to place at the head of the browser page. The system inserts a line to indicate the amount of time the user has been using the system. Next, it inlines the captured and converted JPEG image, placing it within an
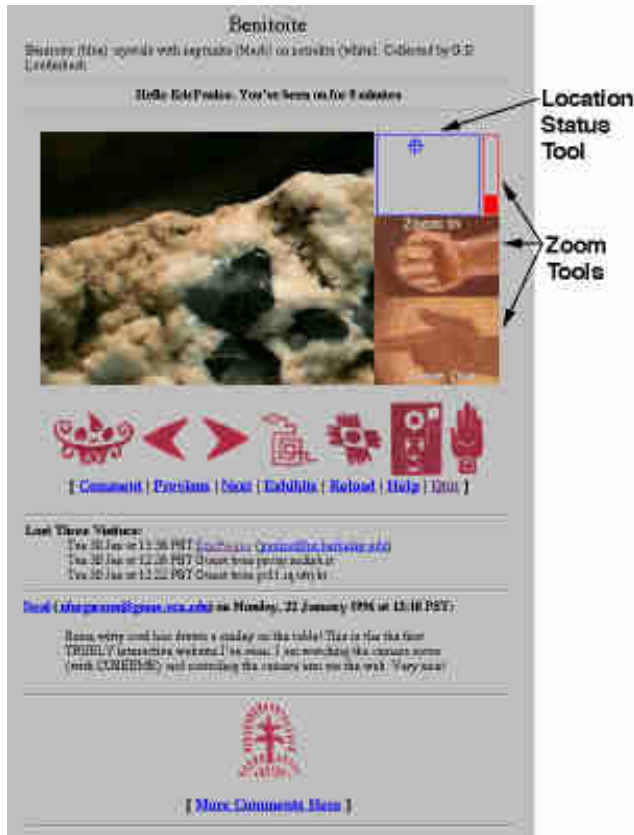
Figure 2: A typical navigation page using the RobotWorld robot

imagemap with a unique randomly assigned number. To the right, various navigational tools are layed out. Additional navigation icons are attached below this. These icons allow users to leave comments about the exhibit, move to the next or previous exhibit, returning to the list of exhibits, obtain help, or return home. To convey a sense of presence of other users, the system then displays of the last three visitors into the system. Finally, the various comments left concerning this exhibit are attached to end of the page, completing the delivery of the HTML file. The CGI script writes out an internal user file using the same randomly generated unique number from above. This file contains the state information, such as the user, position, time, and other information concerning the page and accompanying image just delivered. This number is embedded within the page so that requests originating from this page will reference into this corresponding unique status file. This allows for subsequent requests to make their reference relative to the correct position that the user last viewed. The final result of a remote environment navigation request is a page similar to the one depicted in Figure 2.

7   Navigational Tools

After receiving a navigation page, a user may wish to modify the vantage point of the exhibit and obtain a new image. This modification takes place by employing one or more of the navigational tools presented to the user from that page shown in Figure 2. These tools provide the sensation of exploring a remote space by altering the viewpoint.

One navigation option available to the remote user is to scroll the image. Scrolling moves the camera within the same plane as the current view, captures a new image from that location and delivers it to the remote viewer in a new navigation page. This is accomplished by either selecting a portion of the image for fine motion control or the location status tool for more coarse motion. Fine motion requests bring the selected portion of the image directly into the center of the field of view in the subsequent image while coarse motions move the camera to a particular area within the entire defined exhibition space.

Every exhibit allows a user to zoom in closer to an object for a more detailed inspection, as well as zoom out to achieve a wide angle view. Zooming is accomplished through the zoom navigation tool located on the right size of the image. The camera mimics the motion of the thermometer indicator. Users can also make selections directly on the thermometer to better control the zooming.

When the system is employing the operations of the Intelledex 6DOF robot, the rolling and pitching tools (not pictured in the sample navigation page) are presented to the user. Choosing a point on the roll or pitch tool will cause the camera to roll or pitch depending upon the selection and deliver the resulting image from the new vantage.

8   Future Ideas and Discussion

Mechanical Gaze has proved itself as a useful tool for exploring remote environments. However, these worlds are finite and defined by the limits of the workspace of the robot. We have since been developing a new tool for space browsing on a much larger scale. We realized that it was necessary to deliver a more realistic perception of physical embodiment of the user within the remote space being explored. Such as system must immerse the user in the remote world by providing continuity of motion and user control of that motion. These elements would provide the user the visual cues necessary to stitch together the entire visual experiences into a coherent picture of a building and its occupants. We also wanted to provide the user with the means to communicate and interact with the remote world and its real inhabitants using this new system.

Our answer to this problem is a helium-filled blimp of human proportions, or smaller. The blimp or tele-mobot is

a simple device. It has several motors directly connected to small propellers, and no other moving parts. The complexity of the electronics is comparable with a CB radio or small television receiver. The prototype cost approximately $1000. In quantity, it should be possible to produce blimps that perform better than our prototype for $300. Among the mechatronic devices (printer, scanners, etc) that will be connected to computers in the future, blimps will be among the simplest. On board the blimp is a video camera, microphone, speaker, and radio links. The payload is less than a pound. Our first prototype is lifted by a 6 foot by 3 foot cylindrical balloon. Our second prototype has an 8 ounce payload and is lifted by a 5 foot by 2.5 foot balloon which flies in a vertical pose. It can pass through an open doorway, climb a stairwell or enter an elevator. At the other other end of the radio connections, a PC or UNIX machine with frame grabber hardware and a sound card links the blimp's sensors and actuators to the network.

The pilot, elsewhere on the internet, uses a Java [24] applet, running within any Java supporting browser, to deliver continuous commands to the blimp and receive status information back to be displayed and aid the pilot in navigation. In addition, using existing teleconferencing software such as nv or vic for the Mbone or CU-SeeMe for most PC's, the user can receive live video and audio from the remote space. As the user guides the blimp up or down, left or right, network video from the blimp's camera jogs and stutters on the pilot's screen. The user can also send audio back to the remote space so that actual interactive conversations can take place remotely.

The blimp offers the possibility of a wide range of spontaneous, group interactions. Telephones and teleconferencing are intrusive media. The recipient must interrupt whatever they are doing to answer the call. The interaction is either one-on-one, or within a pre-arranged group. A tele-mobot cruising by a group can overhear the conversation, recognize the group members, and decide if it is appropriate to enter the conversation.

Our aim is not to replace direct human interaction but rather to extend it. It allows for mobility in the 3rd dimension which is a wonderful ability in a building, gallery, or large hall. However, the blimp tele-mobot also offer an experience that is often richer than human experience since the blimp can travel into places and deliver views from perspectives where no human could travel, such as high above a production facility or studio.

A blimp tele-mobot has no problems co-habitating with humans. A collision of a blimp with a fixed obstacle or a person is harmless. Even an out of control blimp poses no real threat to people, while such a statement could not be made for other mobile robots.

Difficult motion planning problems are avoided by being

in the air. While most mobile robots with wheels and tracks struggle with varying floor surfaces, and most cannot handle stairs or even a single step, the blimp tele-mobot simply flies over such problems. The blimp also avoids many of the difficult motion planning problems often encountered by wheeled and tracked mobile robots when they enter a room littered with books, files, or cables on the floor.

Overall we have found that the immersive experience of driving a blimp tele-mobot to be very compelling. The motion is smooth and natural and we hope to have a prototype available soon to the WWW community.

Acknowledgments

References

[1] Tim Berners-Lee, Robert Cailliau, Jean-Francios Groff, and Bernd Pollerman. World-wide web: The information universe. *Electronic Networking:Research, Applications and Policy*, 1(2), Westport CT, Spring 1992.

[2] Mark Pesce, Peter Kennard, and Anthony Parisi. Cyberspace. In *First International WWW Conference*, 1994.

[3] David Gelernter. *Mirror Worlds.* Oxford University Press, 1992.

[4] H. John Durrett. *Color and the computer*. Academic Press, 1987.

[5] C.A. Lynch. The technologies of electronic imaging. *Journal of the american society for informaion science*, pages 578–585, September 1991.

[6] M. Ester. Image quality and viewer perception. In *SIGGRAPH 1990 art show*, pages 51–63, August 1990.

[7] J.L. Kirsch and R.A. Kirsch. Storing art images in intellegent computers. In *Leonardo*, volume 23, pages 99–106, 1990.

[8] Raymond Goertz and R. Thompson. Electronically controlled manipulator. *Nucleonics*, 1954.

[9] A.E.R. Greaves. State of the art in nuclear telerobotic: focus on the man/machine connection. In *Transations of the American Nuclear Society*, 1994.

[10] R. D. Ballard. A last long look at titanic. *National Geographic*, December 1986.

[11] C. Ntuen, E. Park, and S. Kimm. A blackboard architecture for human-machine interface in mining teleoperation. In *Human Computer Interaction*, 1993.

[12] C.R. Weisbin and D. Lavery. Nasa rover and telerobotics technology program. In *IEEE Conference on Robotics and Automation Magazine*, 1994.

[13] P.S. Green, J.W. Hill, J.F. Jensen, and A. Shah. Telepresence surgery. In *IEEE Engineering in Medicine and Biology Magazine*, 1995.

[14] J.V. Draper. Teleoperators for advanced manufacturing: applications and human factors callenges. In *International Journal of Human Factors in Manufacturing*, 1995.

[15] R. S. Mosher. Industrial manipulators. *Scientific American*, 211(4), 1964.

[16] R. Tomovic. On man-machine control. *Automatica*, 5, 1969.

[17] Hans Moravec. *Mind Children. The Future of Robot and Human Intellegence*. Harvard University Press, 1988.

[18] Richard Wallace. A notebook active vision system. *IU Workshop*, February 1996.

[19] K. Goldberg, M. Mascha, S. Gentner, N. Rothenberg, C. Sutter, and Jeff Wiegley. Robot teleoperation via www. In *International Conference on Robotics and Automation*. IEEE, May 1995.

[20] **http://www.telescope.org/**.

[21] **http://telerobot.mech.uwa.edu.au/**.

[22] **http://cwis.usc.edu/dept/garden/**.

[23] Karel Čapek. *R.U.R (Rossum's Universal Robots)*. Doubleday, Page, and Company, 1923.

[24] James Gosling and Henry McGilton. The java language environment: A white paper. Technical report, Sun Microsystem, 1995.