Available online at www.sciencedirect.com
SCIENCE DIRECT®
Information Sciences 174 (2005) 37–53
INFORMATION SCIENCES
AN INTERNATIONAL JOURNAL
www.elsevier.com/locate/ins

Modeling hypermedia-based communication

V.V. Kryssanov [a,*], K. Kakusho [b], E.L. Kuleshov [c], M. Minoh [b]

[a] *Information and Communication Sciences Department, Ritsumeikan University, Kusatsu, Japan*
[b] *Academic Center for Computing and Media Studies, Kyoto University, Kyoto, Japan*
[c] *Computer Systems Department, The Far-Eastern National University, Vladivostok, Russia*

Received 29 December 2003; received in revised form 23 August 2004; accepted 24 August 2004

Abstract

In this article, we explore two approaches to modeling hypermedia-based communication. It is argued that the classical conveyor-tube framework is not applicable to the case of computer- and Internet-mediated communication. We then present a simple but very general system-theoretic model of the communication process, propose its mathematical interpretation, and derive several formulas, which qualitatively and quantitatively accord with data obtained on-line. The devised theoretical results generalize and correct the Zipf–Mandelbrot law and can be used in information system design. At the paper's end, we give some conclusions and draw implications for future work.

Keywords: Communication; Hypermedia; System theory; Zipf–Mandelbrot law

* Corresponding author. Tel./fax: +81 77 561 5936.
E-mail addresses: kvvictor@is.ritsumei.ac.jp, kryssanov@mm.media.kyoto-u.ac.jp (V.V. Kryssanov).

0020-0255/$ - see front matter
doi:10.1016/j.ins.2004.08.006

1. Introduction

Hypermedia is an approach to the development of information systems (see reference [4] for an overview). A hypermedia system stores and manages its information as a collection of nodes, (hyper)links, and scripts. Nodes are multiple representations of different media types—chunks of text, photographs, pictures, sounds, movies, etc. Links implement transitions (by association or connection) between nodes, while scripts, which are generalizations of links, are used to combine and control the diverse nodes in a digital document. Hypermedia-based communication may be thought of as the cognitive processing by a hypermedia system's user that includes selecting, organizing, and integrating information represented over time in the networks of nodes.

Hypermedia-based communication differs from "natural" word-based communication in terms of both structure and dynamics. The latter process is usually defined as a linear, single-channel/medium, receiver-focused, but source-driven transfer of information. In contrast, the former is fragmented, multi-channel/media, network-focused and receiver-driven. Owing to the diversity of deployed representations, one might also claim that the former is a generalization of the latter. It is remarkable that, while these two are the main types of communication on the present Internet, only word-based communication has been receiving somewhat adequate technological support, since the existing interface and networked systems rest on the "conveyor-tube" information-theoretic model and its mathematical interpretations developed in quantitative linguistics half a century ago [21,23,14].

The conveyor-tube model deals with the efficiency of transmitting, as well as coding and decoding of the information communicated from an active source/sender to the target/receiver, and it builds on several assumptions about properties of the communication process. Most interesting for us are the assumptions about the statistical homogeneity of communication (i.e. about a linear discourse of a single speaker or writer) and about the negligences of semantic effects (including those due to context) and redundancy (i.e. before transmission, information is to be efficiently encoded) in communication. The model offers an explanation to several important probabilistic phenomena, including Zipf's "second law" (also known as the discrete Pareto distribution; see extensive bibliography in reference [8]) establishing a relation between frequency and count of words in a word-by-word communication, that are currently employed by search engines, automatic text analysis and indexing tools, and other applications sustaining communication on the Internet. This explanation is, however obviously, not valid or relevant for the case of hypermedia-based communication with its non-linear fragmented discourse and manifold information source-nodes, in effect semantic interdependence-links, and inefficient encoding of information (e.g. due to an overlap of different media types).