# Information Derivatives – A New Way to Examine Information Propagation

**Chirag Shah**
School of Communication & Information (SC&I)
Rutgers, The State University of New Jersey
4 Huntington St, New Brunswick, NJ 08901
chirags@rutgers.edu

## ABSTRACT

An increasing amount of information available online is dynamic in nature. That is, it constantly keeps changing with people consuming and adding their own value to it. We propose to study how such information flows from its original source to other locations through the concept of information derivatives. We treat information as an object and consider the original piece of information as the $0^{th}$ derivative. Any subsequent use of that information produces $1^{st}$ derivative, $2^{nd}$ derivative, and so on. We provide a mathematical formulation of this phenomenon and propose a new framework to study information propagation in online sources, and to understand the dynamic nature of information.

## Categories and Subject Descriptors

H.1.1 [**Models and Principles**]: Systems and Information Theory—*Information theory*

## General Terms

Measurement, Theory.

## Keywords

Information propagation, Information derivatives, Information theory, Social media, Network analysis.

## 1. INTRODUCTION

With the changing ways in which information is stored, retrieved, and shared, it is becoming essential to see information as dynamic and constantly changing with its context and usage. Several works have demonstrated how context could be useful in studying dynamic information (e.g., [8, 11]). Easily creating, sharing, and annotating a piece of information has become largely possible due to the ubiquitous nature of social media sites and services. Examples include YouTube for videos, Flickr for photos, delicious for bookmarks, wordpress for blogs, and Twitter for microblogs.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.
*HCIR 2010*, August 22, 2010, New Brunswick, NJ, USA.
Copyright 2010 ACM x-xxxxx-xxx-x/xx/xxxx…$5.00.

In this paper we look at information as an object [3] and propose a new method for studying its propagation through various channels on the Web. In order to explain this method, we introduce the notion of information derivatives and propose a new framework for studying the dynamic nature of information through its propagation and derivatives.

## 2. Information propagation and derivatives

In the field of library and information science (LIS), information is often viewed as an object [3]. Kaye [7] (p.37) stated: *"Library and information professionals, thus have a perception of information as objective matter, defining it instrumentally in terms of data, facts, knowledge and opinion, to be applied to learning, decision making, problem solving and other tasks.*" Rowley [10] also reaffirmed that LIS professionals tend to treat information as an object and create a system's view of information.

There is, however, also other works in the field that treat information as a vehicle for knowledge. For instance, Hill [6] used the term information to mean recordable knowledge, asserting that knowledge is more than information and that information must be recordable and shareable.

### 2.1 A mathematical view of information derivatives

There are many other views of information, but for the work presented here, we will treat information as a thing, acknowledging that such information is recordable, shareable, and modifiable. We follow Brookes' [2] fundamental equation of information theory:

$$(S) + \Delta I \rightarrow (S + \Delta S) \quad ... \quad (1)$$

Here, $S$ is the knowledge structure that is modified by the information input $\Delta I$ to produce a new knowledge structure $(S+\Delta S)$. Instead of knowledge structure, we will focus on changing information itself. In other words, we propose:

$$I + \Delta V \rightarrow (I + \Delta I) \quad ... \quad (2)$$

Here, $I$ is an original piece of information, $\Delta V$ is the value added to that information, and $(I+\Delta I)$ is the new information created due to this process. An added value changing the given information is a well-studied concept in information life cycle and related works (e.g., [12]).

Let us consider an example. Using social bookmarking service *delicious*, one could collect bookmarks. Each of these bookmarks is a piece of information ($I$) to which one adds tags and notes ($\Delta V$)

to create one's own value-added personal information *(I+ΔI)*. Here, *ΔV* and *ΔI* are person and/or context dependent. Continuing this example, these bookmarks could be shared with others over delicious. Those who subscribe to them could create their own versions of the bookmarks, thus adding *ΔV* to what is already *(I+ΔI)*. Thus, we can generalize Equation (2) as

$$I^j + \Delta V^j \rightarrow I^{j+1} \quad ... \quad (3)$$

Here, *j* represents the order or the derivative rank. *I* with 0th rank ($I^0$) is the original piece of information (0th derivative). With $\Delta V^0$ added, it transforms to $I^1$, which is the 1st order derivative.

Let us consider this notion of information derivatives with another example. Stacy uploads a picture that she took at a birthday party on Facebook (0th derivative), in which Mark tags people. This tagged photo is now a 1st derivative information. Others in Stacy's social network could see this value-added information and start commenting on that photo. This creates 2nd derivative information. This can continue, not only on Facebook, but also on other channels such as blogs, creating new derivatives of the original piece of information.

## 2.2 A graphical view of information derivatives

Information derivatives can be seen as a directed graph,[1] such as the one shown in Figure 1.

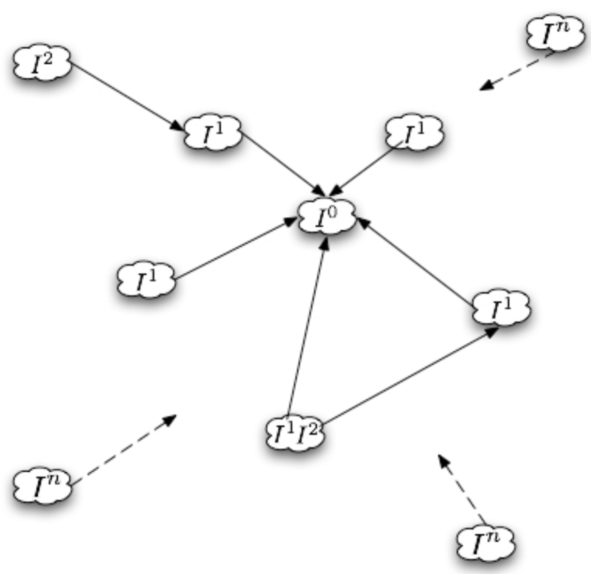

**Figure 1: A graphical view of information derivatives.**

Here, an original piece of information is labeled with $I^0$, which gets propagated through various channels and becomes $I^1$. Note that each of the $I^1$ may be different due to their unique context, usage, and added value. This chain could continue and produce derivatives, $I^2$, $I^3$,..., $I^n$. It is possible that a particular piece of information, for example, $I^jI^k$ is multiple derivatives; that is, it can be produced using derivatives of different lower orders, such as $I^p$ and $I^q$, where $p=j-1$, $q=k-1$, and $p \neq q$. In other words, such information propagation creates many-to-one mapping of information derivatives. In Figure 1, one of the pieces of information is such derivative, labeled with $I^1I^2$.

One way of thinking about the information derivatives from this graphical view is to look at how many hops we need to make from a piece of information to $I^0$. This number indicates the order of that derivative. For instance, for the node labeled with $I^1I^2$, there are two ways to get to $I^0$ – one with one hop and the other with two hops. Thus, this piece of information is 1st order as well as 2nd order derivative with respect to the information represented by $I^0$.

Information propagation through information derivatives can also be represented using other structures, such as tree (Figure 2) and forest, as well. The choice of a structure for portraying and examining information derivatives can depend on the application and the domain in which they are being studied.

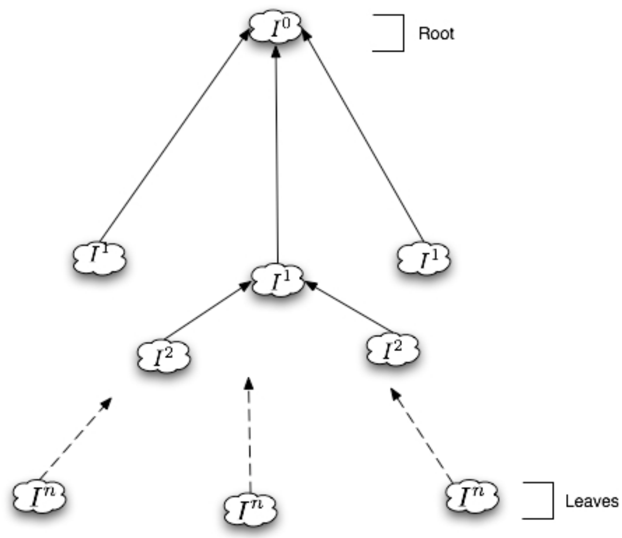


**Figure 2: A tree structure for information derivatives.**

## 2.3 Graph-based characteristics of information derivatives

Various structural views of information derivatives provide us an opportunity to define certain characteristics of them based on the structure as well as how different nodes within this structure are defined and linked. Here we present a few important characteristics that will later be useful for our case study.

### *2.3.1 Diversity or propagation applicability*

A given node in the graphical structure could be used multiple times to create information derivatives. The more a piece of information (node) gets used, the higher its propagation applicability. Diversity of a node is determined by the number of unique derivatives created from that node. Thus, diversity of a derivative node can be defined as:

$$\gamma(I^j) = \sum I^k \quad \text{, where } k = j + 1 \quad ... \quad (4)$$

[1] Note that the direction of the arrows in this representation may seem opposite of the direction for information propagation, but it clarifies the idea of derivatives or links.

Here, all of $I^k$ are derived from $I^j$. Note that these $I^k$ come from unique nodes. In case of the Web, these nodes may be different domains.

This characteristic is similar to in-degree of a node in a graph, often used to measure a node's nature as an authority in network analysis [9].

### 2.3.2 Deriving index

This measures the level of derivation for a piece of information, and is calculated by counting the number of sources it derives. Deriving index is computed as:

$$\theta(I^j) = \sum I^k \quad \text{, where } k = j - 1 \quad ... \quad (5)$$

Here, $I^j$ is derived from $I^k$.

This characteristic is similar to out-degree of a node in a graph, often used to measure a node's nature as a hub in network analysis [9].

### 2.3.3 Density

A piece of information may be derived by a single node several times, possibly by different parts of that node. For instance, if a web-domain is considered as a node, its parts are all the webpages in that domain. Density of an information node is computed by dividing the total number of derivations from that node to the number of unique nodes from which those derivations occur. Thus, density:

$$\delta(I^j) = \frac{\sum I^k}{\sum I_p^k} \quad \text{, where } k = j + 1 \quad ... \quad (6)$$

Here, the denominator is the sum of all the unique nodes (e.g., web-domains) that derive from $I^j$. Note that $\gamma(I^j)$ represents total number of derivations (in-degree) from $I^j$, whereas $\delta(I^j)$ indicates how spread out those incoming links are among unique sources that derive from it.

## 3. DISCUSSION

The ways in which people seek, find, and share information have changed considerably in the recent years, and so are the methods for producing and consuming information online. It is not uncommon for one to discover interesting or relevant information from a friend's blog or a Tweet. In this paper we attempted to formalize such information propagation on the Web. We proposed a notion of information derivatives and presented some of their basic characteristics. Our proposal was inspired by some of the classical works on information as an object, information theory, and network analysis.

The framework presented here using information derivatives can be used for examining almost any case of information flow where information can be treated as an object. One of such domains is blogosphere. Similar to the works of Gruhl et al. [5], and Tremayne et al. [13], we can examine information propagation on certain topics within blogosphere using the notions and properties of information derivatives. Such analysis can help us identify and study connectors (high diversity), and authorities (high deriving index) in a given network.

On the theoretical side, the proposed framework can help us classify and explain data, information, and contextual details, similar to Buckland's [4] discussion about the nature of a document. With the same example of antelope that Buckland gave, antelope itself can be treated as the $0^{th}$ derivative, and the scholarly articles about that antelope can be considered as $1^{st}$ derivatives.[2] One could imagine applying similar definitions to various artifacts considered for preservation or archiving.


## 4. ACKNOWLEDGMENTS

The author is sincerely grateful to Gary Marchionini from UNC Chapel Hill, and Geoffrey Baym from UNC Greensboro for their valuable feedback and criticism of this work. The current article is taken from a larger work, co-authored with Geoffrey Baym, which is currently under review.


## 5. REFERENCES


[1] Briet, S. (1951). *Qu'est-ce que la documentation*. Paris: EDIT.

[2] Brookes, B. C. (1974). Robert Fairthorne and the scope of information science. *Journal of Documentation, 30*(2), 139-152.

[3] Buckland, M. K. (1991). Information as Thing. *Journal of the American Society for Information Science, 42*(5), 351-360.

[4] Buckland, M. K. (1997). What is a "Document"? *Journal of the American Society for Information Science, 48*(9), 804-809.

[5] Gruhl, D., Guha, R., Liben-Nowell, D., & Tomkins, A. (2004). Information diffusion through blogspace. In *Proceedings of World Wide Web (WWW) Conference* (p. 491-501). New York, New York, USA: ACM Press.

[6] Hill, M. W. (1982). Information for innovation: a view from the UK. In *Proceedings of ICSU-AS Information for Innovation*. Amsterdam.

[7] Kaye, D. (1995). The nature of information. *Library Review, 44*(8), 37-48.

[8] Lee, C. (2007). Taking Context Seriously: A Framework for Contextual Information in Digital Collections. UNC SILS TR-2007-04.

[9] Park, H. W., & Thelwall, M. (2003). Hyperlink Analyses of the World Wide Web: A Review. *Journal of Computer-Mediated Communication, 8*(4).

[10] Rowley, J. (1998). What is information? *Information Services & Use, 18*(4), 243-254.

[11] Shah, C. (2009). Mining Contextual Information for Ephemeral Digital Video Preservation. *International Journal of Digital Curation (IJDC), 4*(1), 175-192.

[12] Taylor, R. S. (1982). Value-added processes in the information life cycle. *Journal of the American Society for Information Science, 33*(5), 341-346.

[13] Tremayne, M., Zheng, N., Lee, J. K., & Jeong, J. (2006). Issue publics on the web: Applying network theory to the war blogosphere. *Journal of Computer-Mediated Communication, 12*, 290-310.


[2] Briet [1] (p.7-8) recognized a captured antelope as the primary document, and any article written about it as the secondary documents.