Multimedia情報表現とContent流通System
Multimedia Info-Representation and Content Distribution Systems

第2回
#2

Multimedia情報表現と記述言語 (1) ODA
Multimedia Info-Rep. and Description Language (1) ODA

目次

Contents

ODA標準概要

- Office/Open Document Architecture
  - *事務文書体系*
  - *開放型文書体系*
- ISO/IEC JTC1とITU-Tの共同作業による文書体系の標準化Activityによって標準化される。
- 1980年代当初から標準化活動を開始。
- 1988年に第1段階の標準化活動は終了。
- 1993年頃までに第2段階の標準化活動（HyperODA）は終了。
- 基本的に「成功した標準化方式」ではないが、思想的に学ぶべき事は多い。

Outline of ODA Standard

- Office/Open Document Architecture
- Specified by the joint activities of ISO/IEC JTC1 and ITU-T aiming at the standardization of the document architecture.
- Starting activities in early 1980s.
- Finalizing the 1st step of standardization in 1988.
- Finalizing the 2nd step of standardization in 1993 (HyperODA).
- Not a successful standard (not widely accepted), but lots of design concepts are worth to look through.

時代背景

- ISDN時代の幕開け。
  - *ISDN (Integrated Services Digital Network): 国際標準化された初めてのDigital Subscriber Network*
- Digital Networkを使った様々な情報Mediaの転送と、それらを利用した複合Serviceの提供の兆し。
- 情報交換における互換性の欠如。
  - *世界規模の情報Networkは、Analog電話以外には存在していなかった。*
- Facsimile通信標準化の成功。
- OSI神話の始まり。
  - *OSI: Open Systems Interconnection*

Background of Standardization

- Opening the digital communication world with ISDN.
  - *ISDN (Integrated Services Digital Network): the world first standardized digital subscriber network*
- Realizing various information transmission via digital network, and multimedia information services based on it.
- No interworking and no compatibility on information exchanging.
  - *There was no world-wide network except for the analog telephone network.*
- Great success on standardization of facsimile.
- Start of OSI fanaticism
  - *OSI: Open Systems Interconnection*

Facsimile通信規格

- G1: Analog FAX
- G2: Digital Coding (MH: Modified Huffman) on PSTN
- G3: Digital Coding (MR: Modified Read) on PSTN
- G4: Digital Coding (MMR: Modified MR) on ISDN

- G3の規格化と製品化、及びその爆発的な普及
- 日本の技術的な貢献

- PSTN: Public Switched Telephone Network
- ISDN: Integrated Service Digital Network

Facsimile Standards Suites

- G1: Analog FAX
- G2: Digital Coding (MH: Modified Huffman) on PSTN
- G3: Digital Coding (MR: Modified Read) on PSTN
- G4: Digital Coding (MMR: Modified MR) on ISDN

- Finalizing the specification and successful manufacturing of G3, and its explosive penetration all over the world.
- Lots of technical contributions by Japanese members.

- PSTN: Public Switched Telephone Network
- ISDN: Integrated Service Digital Network

G4規格の詳細

- G4 FAXにおける高度文書交換Service
  - *G4 Class 1: Image transmission only (like G3)*
  - *G4 Class 2: Handling Formatted ODA Docs*
  - *G4 Class 3: Handling Formatted/Processable ODA Docs*

- 上記のClass 2とClass 3に見られる新しいFAXの通信方式を「Mixed Mode通信」と呼んだ。

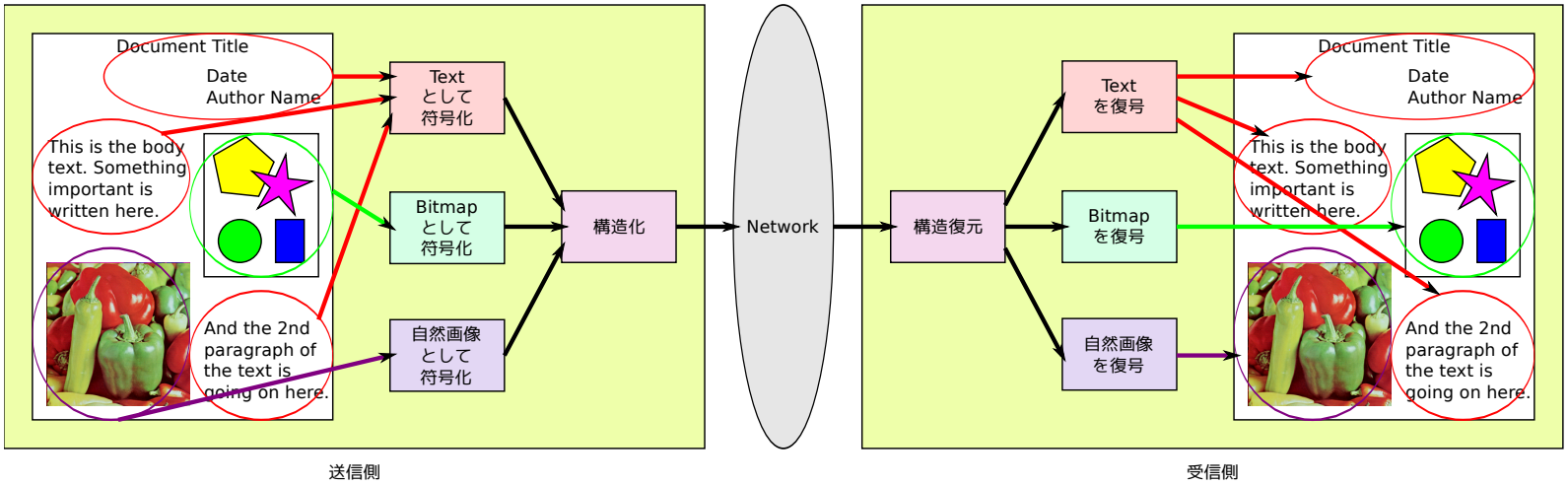
Details of G4 Spec.

- Advanced Document Exchange Service using G4 FAX
  - *G4 Class 1: Image transmission only (like G3)*
  - *G4 Class 2: Handling Formatted ODA Docs*
  - *G4 Class 3: Handling Formatted/Processable ODA Docs*

- New facsimile communication services as introduced by class 2 and 3 were called "Mixed Mode Communication".

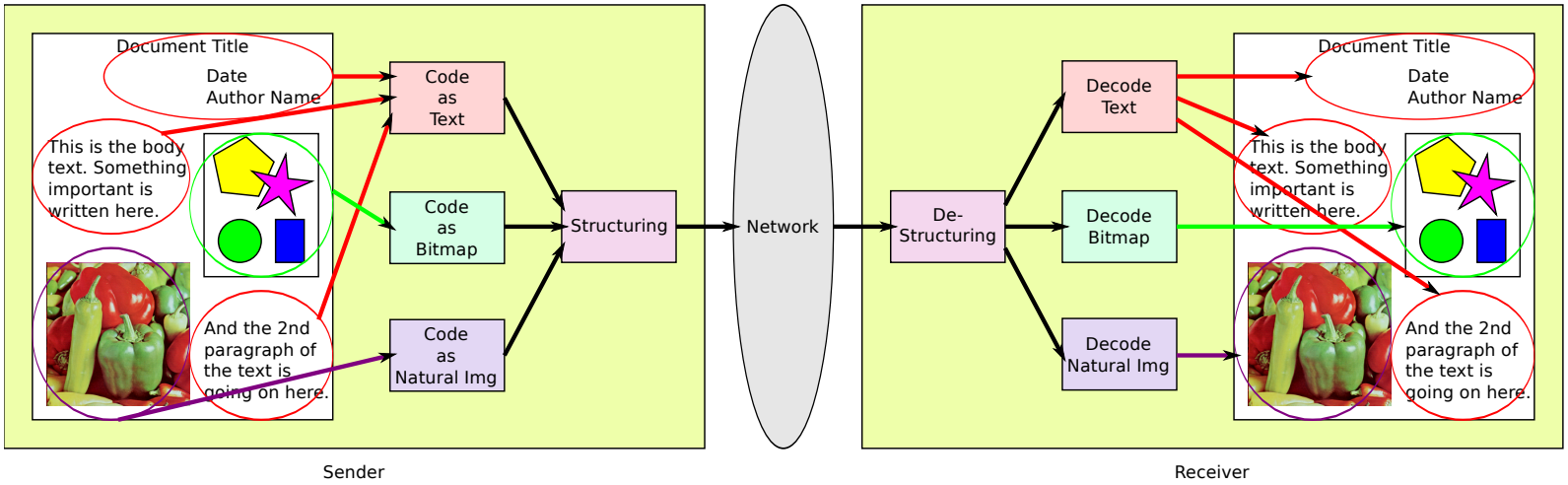
Mixed Mode通信

- 最適な符号化方式を文書中の領域に適応的に割り当てることで、通信効率の劇的な向上が望める。
- 文書の再利用の可能性を導く。
- 更に、DB等を利用した高度文書交換環境を想定した。
- PAX（PC plus FAX）という概念も登場。



Mixed Mode Communication

- Getting drastic high coding efficiency by applying adaptive coding methods to different areas in documents.
- Introducing the possibility of "re-use" of documents.
- Considering much advanced doc-exchanging using DB.
- Introducing a new concept of PAX (PC plus FAX).



OSI (1/2)

- ARPANETの成功（1970年代）
- Internetへの発展（1980年代）
  - *Internetは思わぬ人気を獲得した。*
  - *ARPANET開設当時、Packet交換網の構築に関しAT&Tはじめ世界のTelecom会社は否定的であった。*
- 通信と情報処理の融合が、新しいTelecomの課題として脚光を浴びる。
  - *Computerと通信Digital化の融合 → Packet交換とRouterの発明*
- 通信のProtocol体系・Application体系の再構築の動き。
  - *世界中の研究機関、標準化機関が長年に渡って精力的に活動を続けた。*
  - *M. T. Rose: "The Open Book, A Practical Perspective on OSI", Prentice Hall（邦訳:「実践的OSI論」プレンティスホール・凸版）*
  - *世界的なOSIに対する「熱狂」*

OSI (1/2)

- Success of ARPANET (in 1970s)
- Evolution into Internet (in 1980s)
  - *Internet became popular, unexpectedly.*
  - *Telecom companies, like AT&T, were quite negative against the packet network, at the time of ARPANET starting.*
  - *The convergence of telecom and info-processing was a new subject of telecom industry to be investigated.*
  - *Convergence of computer and digital network -> invention of packet switching and router*
- Reconsidering all the protocols and applications.
  - *Lots of researchers world wide put all their efforts into the OSI.*
  - *Ref: "The Open Book, A Practical Perspective on OSI" by M. T. Rose published by Prentice Hall.*
  - *People were so enthusiastic over OSI.*

OSI (2/2)

- しかしながら、世の中に広まることはなかった。
- いくつかの原因が考察されている。
  - *PCの急速な性能向上と低廉化*
  - *網と端末の役割の変化*

- しかし遺産も多い
  - *ASN.1*
  - *OSI Protocol Model (7 Layer Model)*
  - *下位Layerにおける様々なProtocol*

OSI (2/2)

- Eventually, it was not widely accepted and implemented.
- Some reasons were mentioned.
  - *explosive evolution of PC in terms of processing speed and low price.*
  - *changing roles of network and terminal.*

- However, we are still using some legacies of OSI
  - *ASN.1*
  - *OSI protocol model (7 layer model)*
  - *Various protocols in lower layers*

ODA Standards Suite (1/2)

ISO/IEC 8613, Information technology -- Open Document Architecture (ODA) and Interchange Format

- ISO/IEC 8613-1:1994, Introduction and general principles
  - *ISO/IEC 8613-1:1994/Cor 1:1998*
- ISO/IEC 8613-2:1995, Document structures
  - *ISO/IEC 8613-2:1995/Cor 1:1998*
  - *ISO/IEC 8613-2:1995/Cor 2:1998*
- ISO/IEC 8613-3:1995, Abstract interface for the manipulation of ODA documents
- ISO/IEC 8613-4:1994, Document profile
  - *ISO/IEC 8613-4:1994/Cor 1:1998*
  - *ISO/IEC 8613-4:1994/Cor 2:1998*
- ISO/IEC 8613-5:1994, Open Document Interchange Format
  - *ISO/IEC 8613-5:1994/Cor 1:1998*
  - *ISO/IEC 8613-5:1994/Cor 2:1998*

ODA Standards Suite (2/2)

- ISO/IEC 8613-6:1994, Character content architectures
  - *ISO/IEC 8613-6:1994/Cor 1:1998*
- ISO/IEC 8613-7:1994, Raster graphics content architectures
  - *ISO/IEC 8613-7:1994/Cor 1:1998*
  - *ISO/IEC 8613-7:1994/Amd 1:1998*
- ISO/IEC 8613-8:1994, Geometric graphics content
- ISO/IEC 8613-9:1996, Audio content architectures
- ISO/IEC 8613-10:1995, Formal specifications
- ISO/IEC 8613-11:1995, Tabular structures and tabular
- ISO/IEC 8613-12:1996, Identification of document
- ISO/IEC 8613-14:1997, Temporal relationships and non-linear structures
- ISO/IEC TR 10183-1:1993, Technical Report on ISO 8613 implementation testing -- Part 1: Testing methodology
- ISO/IEC TR 10183-2:1993, Technical Report on ISO 8613 implementation testing -- Part 2: Framework for abstract test cases

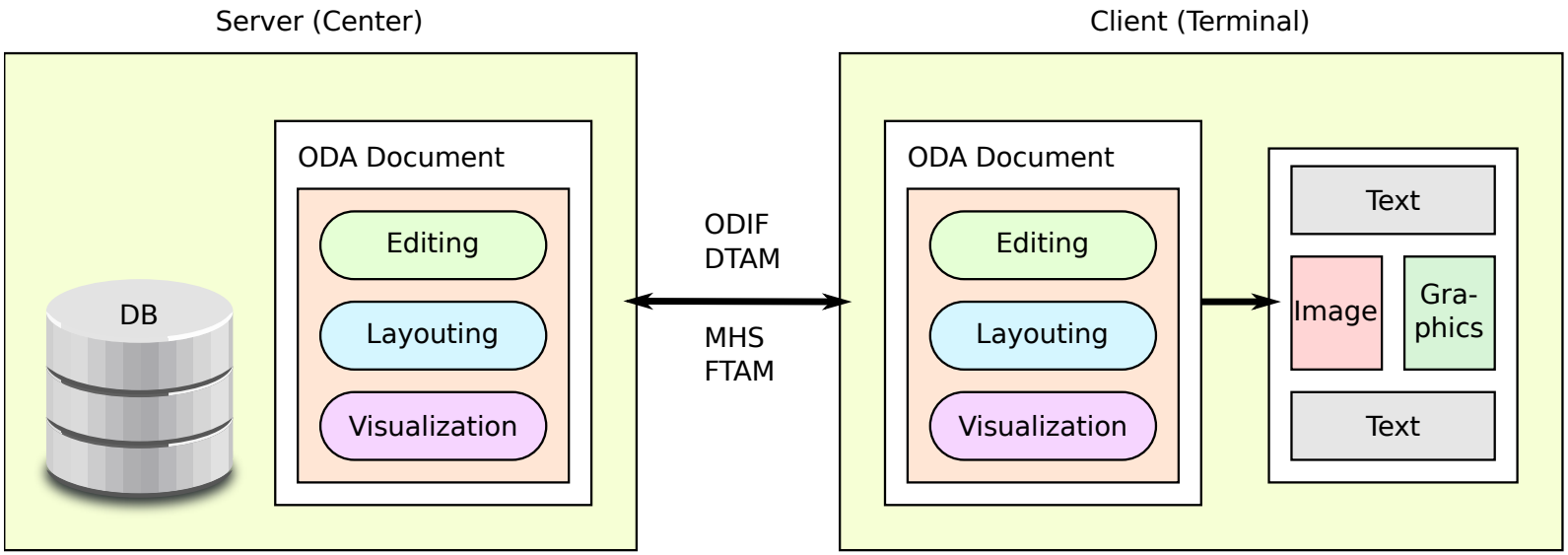
ODAの基本思想

- 文書の互換性の確保
  - *Multimedia通信規格の整備*
  - *文書の構造化方式の標準化*
  - *文書の符号化方式の標準化*
- 文書の再利用
  - *受信した文章の容易な改編*
- DBと連動した文書の利用環境
  - *Templateを利用した効率的な文書作成*
  - *Templateの共通化による効率的な文書交換*
- これら高度文書交換を支える各種のOSI Protocolの利用と開発

Basic Concepts of ODA

- Ensuring the document portability and exchangeability
  - *Providing multimedia communication protocols*
  - *Standardizing document structuring method*
  - *Standardizing document coding scheme*
- Promoting re-using of documents
- Promoting the document processing environment supported by DB
  - *Efficient document creation using template*
  - *Efficient document exchange by common template*
- Developing various OSI protocol suites to support above-mentioned advanced document exchange

ODA System Concept



- ODIF: Open Document Interchange Format
- DTAM: Document Transfer and Manipulation
- MHS: Message Handling System
- FTAM: File Transfer and Manipulation

どのような構造が文書にあるか?

N-020

A Proposal of Contents Distribution Model and Analysis of Illegal Contents Distribution on the Small-World Network

Pao Sriprasertsuk† Wataru Kameyama††

1. Introduction

In general, there are two kinds of contents flows in an information distribution; one is the primary information distribution, and the other is the secondary information distribution. The primary information distribution is the distribution done by providers or broadcasters to consumers through certain kinds of media such as television, newspaper, etc. The secondary information distribution is the distribution done by users to users. For examples, one student copies rental DVD and gives to his or her friends. Today, the advanced information technologies enable the secondary information distribution to be able to perform by various methods and media.

Nevertheless, the secondary information distribution has been also increasing the power of illegal distribution of contents, and contents industries have been suffering from it, particularly by peer-to-peer file sharing (p2p) networks. The study of [1] has predicted that global peer-to-peer networks will be effectively stopped by legal means.

In addition, there is another way to illegally distribute contents that is the Small-World Network(SWN)[2] or social network. With the advanced information communication technologies and availability of cheap media, storage devices and high bandwidth networks, distributing contents in the SWN has been becoming easier and more powerful than before. Therefore, it is necessary to analyze and control the power of secondary information distribution.

We have investigated how the secondary information distribution affects the information circulation based on the statistical observation[3]. However, the proposed model does not consider the SWN structure but only the statistical aspects of information distribution being consumed. There are researches[4][5][6] using the SIR model[7] to analyze the effectiveness of information distribution in the SWN. We consider the SIR model is not suitable to represent the real world information propagation cycle since it does not fully take into account the nature of human behavior for information distribution.

In this paper, we propose an information distribution model and equations considering the human behavior. Subsequently, the proposed model is simulated and analyzed by applying the SWN and it characteristics to analyze illegal contents distribution.

2. The Proposed Information Distribution Model

2.1 Model Overview

With regard to the human behavior state for information distribution, we consider that there are three states that are "Unknown", "Known" and "Distribute". The proposed model is shown in Figure 1, and the definitions are described as below.

- Unknown State (U)
  Individuals in this state do not know information. They either do not receive information yet or they forget it. In the case of forgetting information, individuals transit from Known State to this state.
- Known State (K)
  In this state, individuals know information but do not have any action to the information distribution.
- Distribute State (D)
  Individuals in this state are active to distribute information. The distribution by their own intentions and other individuals requests are considered as the same.
- Probability of Becoming Known State (B_1)
  This parameter shows the probability of individuals in Unknown State to change into Known State. For example, some individuals in Unknown State are informed with the information by individuals in Distribute State.
- Probability of Becoming Distribute State (B_2)
  This parameter shows the probability of individuals in Known State to change into Distribute State. For example, if individuals in Know State start to distribute the information, they move from Known State to Distribute State.
- Probability of Returning to Unknown State (R_1)
  This parameter shows the probability of individuals in Known State to change into Unknown State. For example, they forget the information because they are not interested or the information becomes stale after time passes.
- Probability of Returning to Known State (R_2)
  This parameter shows the probability of individuals in Distribute State to change into Known State. For example, after individuals distributed the information, they may change their mind to stop the action. Thus, their state is changed to Known State.

With regards to Figure 1 and the above definitions, the notations with time series are defined as below.

- N : the number of all individuals in the network
- $U(t)$: the number of Unknown State individuals at time t
- $K(t)$: the number of Known State individuals at time t
- $D(t)$: the number of Distribute State individuals at time t

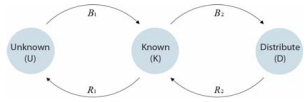


Figure 1: The Proposed Model

2.2 The Process of Transiting to Known State from Unknown State

New individuals get information by being in contact with individuals in Distribute State. We assume that each individual in Distribute State contacts k neighbors in each period, and the probability of successful distribution of each individual in Distribute State is B_1. Therefore, the probability of successful distribution for individuals in Unknown State depends on numbers of neighbors in Distribute State. This probability is defined as Equation 1, where G_i is successful distribution probability of individual i and n is the number of neighbors of it in Distribute State.

$$G_i = 1-(1-B_1)^n \tag{1}$$

2.3 Dynamism in R_1

We apply the forgetting curve theory of Hermann Ebbinghaus[8] because it has clearly stood the test of time and has been validated in many numbers of dissertations and follow-up studies. As discussed in section 2.2, new individuals might transit from UnKnown and Distribute State to Known State, and individuals might leave Known State any time. We consider that, when those new individuals transit to Known State, they start to forget the information with the probability of R_1. Thus, $R_1(t)_i$ for individual i is defined in Equation2 where t_p is time when individual i transit to Know State and S is the relative strength of memory.

$$R_1(t)_i = 1-e^{-(\frac{t-t_p}{S})} \tag{2}$$

2.4 Dynamism in B_2 and R_2

We also consider B_2 and R_2 as dynamic parameters in this paper because their values depend on motivation and feeling of individuals to information itself. As we know, human intention, motivation and feeling of information distribution are so varied due to information itself. Such variations are able to be represented in many kinds of graph such as asymmetric bell curve. Hence, we propose Equation 3 to represent $B_2(t)$ and $R_2(t)$. Equation 3 is flexible to produce various kind of graphs by changing μ, λ_{up}, λ_{down} and β. Changing μ, λ_{up}, λ_{down} and β means changing time when the graph reach maximum value, changing graph scale when $t \le \mu$, changing graph scale when $t > \mu$, and changing maximum value of y-axis, respectively.

$$B_2(t), R_2(t) = \begin{cases} \dfrac{e^{-(t-\mu)^2\times\lambda_{up}}}{\beta} & (t \le \mu) \\ \dfrac{e^{-(t-\mu)^2\times\lambda_{down}}}{\beta} & (t > \mu) \end{cases} \tag{3}$$

4. Simulation and Analytical Results

4.1 Simulation Overview

In this paper, we use the Newman-Watts(NW) SWN[9] instead of the Watt-Strogatz(WS) SWN, because there is probability for the WS model to be broken into unconnected cluster and the average distance between pairs of nodes on the graph is poorly defined due to the rewiring connection. The parameters for the simulation are shown in Table 2 where $K(0)$ and k are number of initial Known State individuals at the beginning of information distribution and the average number of connections of each node in the network, respectively.

Table 2: A Default Parameter Values

Parameter	Value
B_1	0.5
$R_1(t)$	Equation 2 with S = 100
$B_2(t)$	Equation 3 with μ = 5, λ_{up}=0.1, λ_{down}=0.005, β = 1
$R_2(t)$	Equation 3 with μ = 450, λ_{up}=0.008, λ_{down}=0, β = 1
$K(0)$	10
k	4
N	10,000
rang of t	1 to 500
shortcut	50

$$DER = \frac{1}{T\cdot N}\sum_{t=1}^{T} K(t)+D(t) \tag{4}$$

In order to observe the impact of distribution effectiveness, we define a parameter calling Distribution Effectiveness Rate(DER) and its equation is shown in Equation 4 whereas T is rang of t. We define the nodes in the network which have ability to copy and redistribute contents as Bad nodes, and nodes which don't have ability to redistribute contents are Constraint nodes. Firstly, we suppose that all nodes in the network are Bad nodes. Subsequently, we conduct simulations by changing Bad nodes to be Constraint nodes at constant rate and observing impact to DER.

4.2 Simulation Results and Considerations

Having conducted the simulations based on one-dimensional ring lattice SWN and two-dimensional square lattice SWN, the results are shown in Figure 2 and Figure 3 whereas y-axis and x-axis are DER and number of Constraint nodes, respectively.

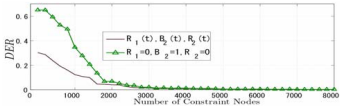


Figure 2: The Impact of Number of Constraint Nodes to DER in One-Dimensional Ring Lattice SWN.

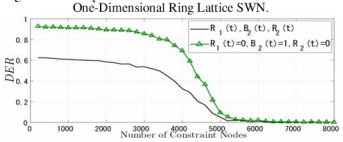


Figure 3: The Impact of Number of Constraint Nodes to DER in Two-Dimensional Square Lattice SWN.

As seen in Figure 2 and 3, the line graph shows the impact when we use $R_1(t)$, $B_2(t)$, $R_2(t)$ from Table 2, and triangle line graph shows the impact when we use $R_1(t) = 0$, $B_2 = 1$ and $R_2 = 0$ as static values. The results show DER is rapidly decreasing until Constraint nodes approximately reach 2,000 in Figure 2 and 5,000 in Figure 3. Surprisingly, the results imply that illegal contents distribution in the network which has 4 average degree distribution and a few of shortcuts can be controlled by suppressing the number of Bad nodes less than 80 percent for one-dimensional SWN and 50 percent for two-dimensional SWN, even though all nodes in the network have motivation to redistribute the contents(triangle ling graph).

The results also show that two-dimensional square lattice SWN is more appropriate to represent and analyze real world content distribution on SWN. Because only 20 percent of Constraint nodes in one-dimensional SWN can stop effectively illegal content distribution, and this kind of phenomenon is rare in the real world.

5. Future Works

There are remaining issues in our work. For examples, we will estimate the appropriate values of parameters in the proposed model for the real world social network. Furthermore, some influencing factors and their impacts for real-world contents distribution will be investigated and analyzed by using the proposed model to analyze and control the illegal contents distribution.

References
[1] P. Biddle, P. England, M. Peinado, and B.Willman: The Darknet and the Future of Content Distribution, In Proceedings of the 2002 ACM Workshop on Digital Rights Management, (2002).
[2] D.J.Watts and S.H. Strogatz: Collective dynamics of 'small world' networks, Nature, vol. 393, pp. 440-442,(1998).
[3] S. Pao, S. Akiko and K. Wataru: On Content Distribution Model and Analyzing Distribution Effectiveness, Journal of Information Processing Society Japan, vol.48 , pp. 2878-2891, No.8, (2007).
[4] H. Liang, P. Kwangho and L. Ying-Chang: Information propagation on modular networks, Physical Review E, vol. 73, 035103(R), (2006).
[5] F. Feng, L. Lianghuan, W. Long: Information propagation in a novel hierarchical network, Arxiv:math/0605293 , (2006).
[6] M. Yamir, Maziar, F.P. Amalio: Dynamics of rumor spreading in complex networks, Physical Review E, vol. 69, 066130, (2004).
[7] W.O Kermack and A.G. Mckendrick: A Contribution to the Mathematical Theory of Epidemics, Proceedings of the Royal Society of London A115, pp. 700-721, (1927).
[8] H. Ebbinghaus: Memory: A Contribution to Experimental Psychology, Columbia University Press, New York, (1913).
[9] M.E.J. Newman and D.J.Watts: Renormalization group analysis of the SWN model, Physics Letters A, vol. 263, pp. 341-346, (1999).

† Global Information and Telecommunication Institute, Waseda University
†† Graduate School of Global Information and Telecommunication Studies, Waseda University

FIT2008（第7回情報科学技術フォーラム）

What Structure in Document?

N-020

A Proposal of Contents Distribution Model and Analysis of Illegal Contents Distribution on the Small-World Network

Pao Sriprasertsuk† Wataru Kameyama††

1. Introduction

In general, there are two kinds of contents flows in an information distribution; one is the primary information distribution, and the other is the secondary information distribution. The primary information distribution is the distribution done by providers or broadcasters to consumers through certain kinds of media such as television, newspaper, etc. The secondary information distribution is the distribution done by users to users. For examples, one student copies rental DVD and gives to his or her friends. Today, the advanced information technologies enable the secondary information distribution to be able to perform by various methods and media.

Nevertheless, the secondary information distribution has been also increasing the power of illegal distribution of contents, and contents industries have been suffering from it, particularly by peer-to-peer file sharing (p2p) networks. The study of [1] has predicted that global peer-to-peer networks will be effectively stopped by legal means.

In addition, there is another way to illegally distribute contents that is the Small-World Network(SWN)[2] or social network. With the advanced information communication technologies and availability of cheap media, storage devices and high bandwidth networks, distributing contents in the SWN has been becoming easier and more powerful than before. Therefore, it is necessary to analyze and control the power of secondary information distribution.

We have investigated how the secondary information distribution affects the information circulation based on the statistical observation[3]. However, the proposed model does not consider the SWN structure but only the statistical aspects of information distribution being consumed. There are researches[4][5][6] using the SIR model[7] to analyze the effectiveness of information distribution in the SWN. We consider the SIR model is not suitable to represent the real world information propagation cycle since it does not fully take into account the nature of human behavior for information distribution.

In this paper, we propose an information distribution model and equations considering the human behavior. Subsequently, the proposed model is simulated and analyzed by applying the SWN and it characteristics to analyze illegal contents distribution.

2. The Proposed Information Distribution Model

2.1 Model Overview

With regard to the human behavior state for information distribution, we consider that there are three states that are "Unknown", "Known" and "Distribute". The proposed model is shown in Figure 1, and the definitions are described as below.

- Unknown State (U)
  Individuals in this state do not know information. They either do not receive information yet or they forget it. In the case of forgetting information, individuals transit from Known State to this state.
- Known State (K)
  In this state, individuals know information but do not have any action to the information distribution.
- Distribute State (D)
  Individuals in this state are active to distribute information. The distribution by their own intentions and other individuals requests are considered as the same.
- Probability of Becoming Known State (B_1)
  This parameter shows the probability of individuals in Unknown State to change into Known State. For example, some individuals in Unknown State are informed with the information by individuals in Distribute State.
- Probability of Becoming Distribute State (B_2)
  This parameter shows the probability of individuals in Known State to change into Distribute State. For example, if individuals in Know State start to distribute the information, they move from Known State to Distribute State.
- Probability of Returning to Unknown State (R_1)
  This parameter shows the probability of individuals in Known State to change into Unknown State. For example, they forget the information because they are not interested or the information becomes stale after time passes.
- Probability of Returning to Known State (R_2)
  This parameter shows the probability of individuals in Distribute State to change into Known State. For example, after individuals distributed the information, they may change their mind to stop the action. Thus, their state is changed to Known State.

With regards to Figure 1 and the above definitions, the notations with time series are defined as below.

- N : the number of all individuals in the network
- $U(t)$: the number of Unknown State individuals at time t
- $K(t)$: the number of Known State individuals at time t
- $D(t)$: the number of Distribute State individuals at time t

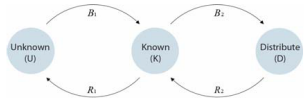


Figure 1: The Proposed Model

2.2 The Process of Transiting to Known State from Unknown State

New individuals get information by being in contact with individuals in Distribute State. We assume that each individual in Distribute State contacts k neighbors in each period, and the probability of successful distribution of each individual in Distribute State is B_1. Therefore, the probability of successful distribution for individuals in Unknown State depends on numbers of neighbors in Distribute State. This probability is defined as Equation 1, where G_i is successful distribution probability of individual i and n is the number of neighbors of it in Distribute State.

$$G_i = 1-(1-B_1)^n \tag{1}$$

2.3 Dynamism in R_1

We apply the forgetting curve theory of Hermann Ebbinghaus[8] because it has clearly stood the test of time and has been validated in many numbers of dissertations and follow-up studies. As discussed in section 2.2, new individuals might transit from UnKnown and Distribute State to Known State, and individuals might leave Known State any time. We consider that, when those new individuals transit to Known State, they start to forget the information with the probability of R_1. Thus, $R_1(t)_i$ for individual i is defined in Equation2 where t_p is time when individual i transit to Know State and S is the relative strength of memory.

$$R_1(t)_i = 1-e^{-(\frac{t-t_p}{S})} \tag{2}$$

2.4 Dynamism in B_2 and R_2

We also consider B_2 and R_2 as dynamic parameters in this paper because their values depend on motivation and feeling of individuals to information itself. As we know, human intention, motivation and feeling of information distribution are so varied due to information itself. Such variations are able to be represented in many kinds of graph such as asymmetric bell curve. Hence, we propose Equation 3 to represent $B_2(t)$ and $R_2(t)$. Equation 3 is flexible to produce various kind of graphs by changing μ, λ_{up}, λ_{down} and β. Changing μ, λ_{up}, λ_{down} and β means changing time when the graph reach maximum value, changing graph scale when $t \le \mu$, changing graph scale when $t > \mu$, and changing maximum value of y-axis, respectively.

$$B_2(t), R_2(t) = \begin{cases} \dfrac{e^{-(t-\mu)^2\times\lambda_{up}}}{\beta} & (t \le \mu) \\ \dfrac{e^{-(t-\mu)^2\times\lambda_{down}}}{\beta} & (t > \mu) \end{cases} \tag{3}$$

4. Simulation and Analytical Results

4.1 Simulation Overview

In this paper, we use the Newman-Watts(NW) SWN[9] instead of the Watt-Strogatz(WS) SWN, because there is probability for the WS model to be broken into unconnected cluster and the average distance between pairs of nodes on the graph is poorly defined due to the rewiring connection. The parameters for the simulation are shown in Table 2 where $K(0)$ and k are number of initial Known State individuals at the beginning of information distribution and the average number of connections of each node in the network, respectively.

Table 2: A Default Parameter Values

Parameter	Value
B_1	0.5
$R_1(t)$	Equation 2 with S = 100
$B_2(t)$	Equation 3 with μ = 5, λ_{up}=0.1, λ_{down}=0.005, β = 1
$R_2(t)$	Equation 3 with μ = 450, λ_{up}=0.008, λ_{down}=0, β = 1
$K(0)$	10
k	4
N	10,000
rang of t	1 to 500
shortcut	50

$$DER = \frac{1}{T\cdot N}\sum_{t=1}^{T} K(t)+D(t) \tag{4}$$

In order to observe the impact of distribution effectiveness, we define a parameter calling Distribution Effectiveness Rate(DER) and its equation is shown in Equation 4 whereas T is rang of t. We define the nodes in the network which have ability to copy and redistribute contents as Bad nodes, and nodes which don't have ability to redistribute contents are Constraint nodes. Firstly, we suppose that all nodes in the network are Bad nodes. Subsequently, we conduct simulations by changing Bad nodes to be Constraint nodes at constant rate and observing impact to DER.

4.2 Simulation Results and Considerations

Having conducted the simulations based on one-dimensional ring lattice SWN and two-dimensional square lattice SWN, the results are shown in Figure 2 and Figure 3 whereas y-axis and x-axis are DER and number of Constraint nodes, respectively.

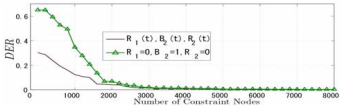


Figure 2: The Impact of Number of Constraint Nodes to DER in One-Dimensional Ring Lattice SWN.

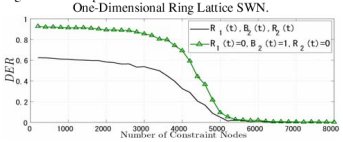


Figure 3: The Impact of Number of Constraint Nodes to DER in Two-Dimensional Square Lattice SWN.

As seen in Figure 2 and 3, the line graph shows the impact when we use $R_1(t)$, $B_2(t)$, $R_2(t)$ from Table 2, and triangle line graph shows the impact when we use $R_1(t) = 0$, $B_2 = 1$ and $R_2 = 0$ as static values. The results show DER is rapidly decreasing until Constraint nodes approximately reach 2,000 in Figure 2 and 5,000 in Figure 3. Surprisingly, the results imply that illegal contents distribution in the network which has 4 average degree distribution and a few of shortcuts can be controlled by suppressing the number of Bad nodes less than 80 percent for one-dimensional SWN and 50 percent for two-dimensional SWN, even though all nodes in the network have motivation to redistribute the contents(triangle ling graph).

The results also show that two-dimensional square lattice SWN is more appropriate to represent and analyze real world content distribution on SWN. Because only 20 percent of Constraint nodes in one-dimensional SWN can stop effectively illegal content distribution, and this kind of phenomenon is rare in the real world.

5. Future Works

There are remaining issues in our work. For examples, we will estimate the appropriate values of parameters in the proposed model for the real world social network. Furthermore, some influencing factors and their impacts for real-world contents distribution will be investigated and analyzed by using the proposed model to analyze and control the illegal contents distribution.

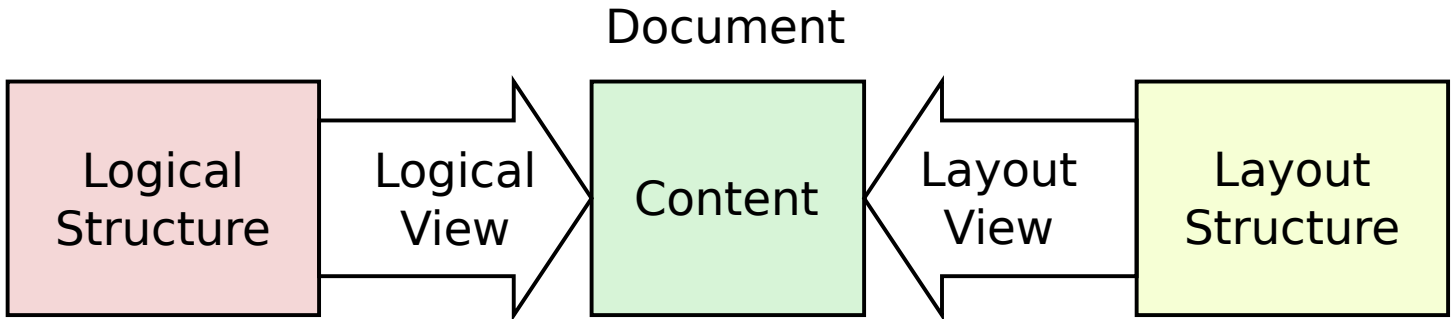
References
[1] P. Biddle, P. England, M. Peinado, and B.Willman: The Darknet and the Future of Content Distribution, In Proceedings of the 2002 ACM Workshop on Digital Rights Management, (2002).
[2] D.J.Watts and S.H. Strogatz: Collective dynamics of 'small world' networks, Nature, vol. 393, pp. 440-442,(1998).
[3] S. Pao, S. Akiko and K. Wataru: On Content Distribution Model and Analyzing Distribution Effectiveness, Journal of Information Processing Society Japan, vol.48 , pp. 2878-2891, No.8, (2007).
[4] H. Liang, P. Kwangho and L. Ying-Chang: Information propagation on modular networks, Physical Review E, vol. 73, 035103(R), (2006).
[5] F. Feng, L. Lianghuan, W. Long: Information propagation in a novel hierarchical network, Arxiv:math/0605293 , (2006).
[6] M. Yamir, Maziar, F.P. Amalio: Dynamics of rumor spreading in complex networks, Physical Review E, vol. 69, 066130, (2004).
[7] W.O Kermack and A.G. Mckendrick: A Contribution to the Mathematical Theory of Epidemics, Proceedings of the Royal Society of London A115, pp. 700-721, (1927).
[8] H. Ebbinghaus: Memory: A Contribution to Experimental Psychology, Columbia University Press, New York, (1913).
[9] M.E.J. Newman and D.J.Watts: Renormalization group analysis of the SWN model, Physics Letters A, vol. 263, pp. 341-346, (1999).

† Global Information and Telecommunication Institute, Waseda University
†† Graduate School of Global Information and Telecommunication Studies, Waseda University

ODAのDocument Data表現

- 文書には2つの構造がある。
  - *表題・章・節という「文書の意味を表す構造」*
  - *文や図を紙またはDisplayのどこに配置するかという「割りつけを表す構造」*
- 前者を「Logical Structure」、後者を「Layout Structure」と呼ぶ。
- しかもこの2つの構造は独立な情報である。
- つまり、文書には、「Logical View」と「Layout View」という2面性が存在する。

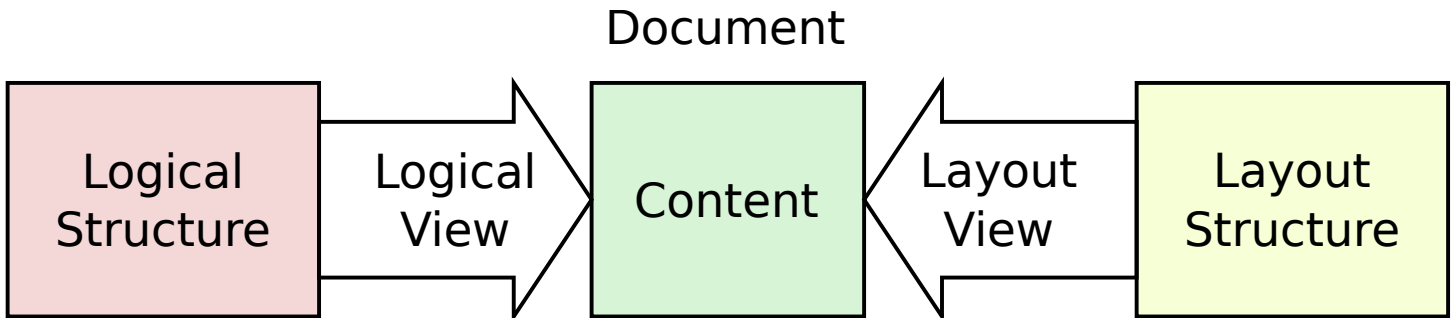
- ODAでは、この2つの構造を用いて「Documentという情報」を構造化、符号化した。
- 各構造の最終的な「中身」はBlockと呼ばれ、Content Architecture によって表されるとした。



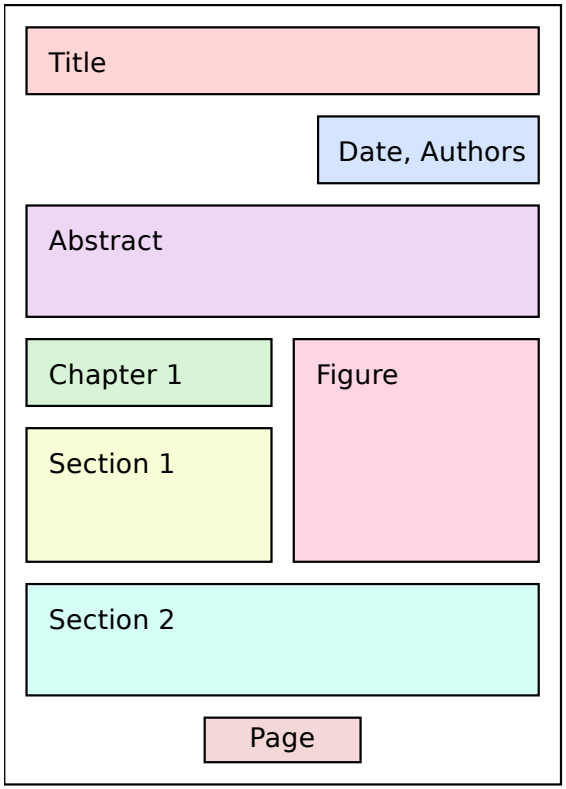
Document Data Representation in ODA

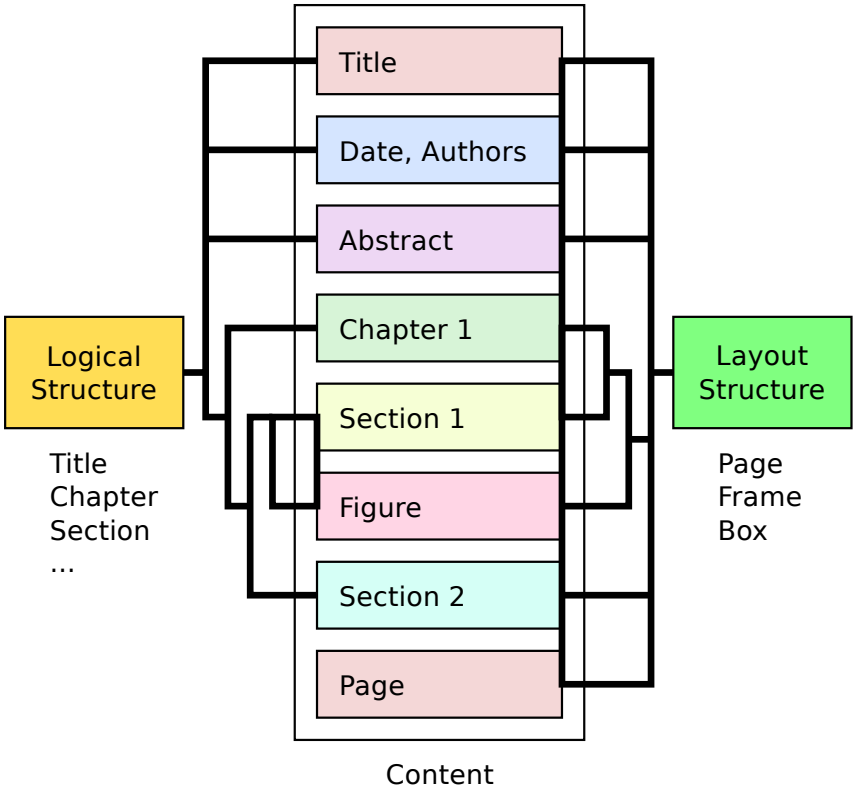
- In a document, two structures are to be observed:
  - *Structure representing the semantics of document, such as "title", "chapter" and "section".*
  - *Structure representing the layout of document, such as where to place sentences and figures on paper or display.*
- The former is called "Logical Structure", and the latter is called "Layout Structure".
- These structures (or info) are independent each other.
- In other words, there are two types of views in a document, "Logical View" and "Layout View".

- In ODA, information as in documents is structured and coded using above two structures.
- The elements (or leaves) of those structures are called "Block" containing contents, represented by "Content Architecture".

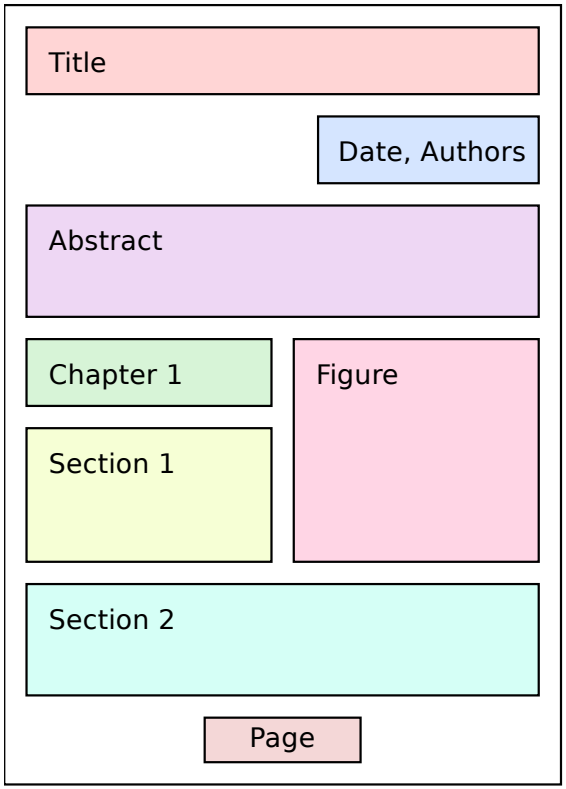


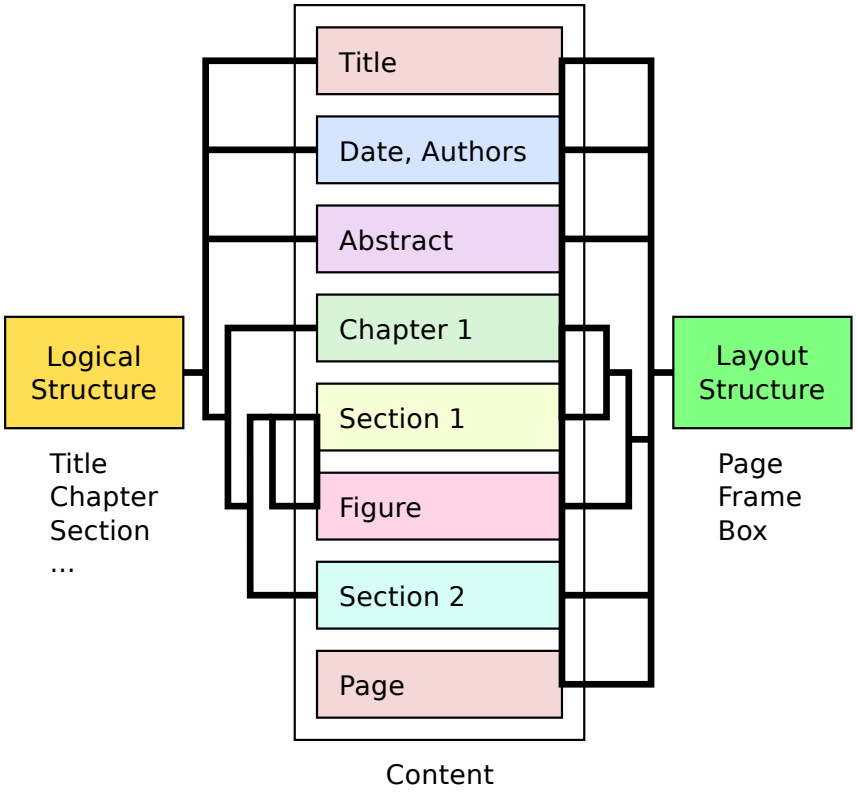
ODAによる文書構造


Title
Date, Authors
Abstract
Chapter 1
Figure
Section 1
Section 2
Page


Logical Structure
Title
Chapter
Section
...
Title
Date, Authors
Abstract
Chapter 1
Section 1
Figure
Section 2
Page
Content
Layout Structure
Page
Frame
Box

Document Structure by ODA





Logical/Layout Structure, Content Architecture

- Logical Structure
  - *文書に固有な「論理構造」を抽出し、文書毎にInstantiation したもの。*
  - *論理構造は文書毎に異なる。*
  - *Layoutに全く依存しない、文書の本質的な構造といえる。*
- Layout Structure
  - *Page, Frame, Blockという構造を持つ。*
  - *Pageは任意数のFrameとBlockを包含できる。*
  - *Frameは任意数のBlockを包含できる。*
  - *Blockはそれ以上分割できず、Content Architectureに沿った「内容」を含む。*
- Content Architecture
  - *Blockには単一の符号化規則で表される情報内容が付属する。*
    - Character Content Architecture (文字)
    - Raster Graphics Content Architecture (Facsimile画像)
    - Geometric Graphics Content (CGMを使用したCG)
    - Audio Content Architecture (音)
    - …

Logical/Layout Structures, Content Architecture

- Logical Structure
  - *A structured info peculiar to a certain type of documents extracted from a document, and instantiated for a document.*
  - *Logical structures are different if the type of documents is different.*
  - *It's independent from layout, i.e. essential structure of documents.*
- Layout Structure
  - *It has a structure of page, frame, and block.*
  - *A page may contain any number of frames and blocks.*
  - *A frame may contain any number of blocks.*
  - *A block may not be divided and may contain "contents" represented by "Content Architecture".*
- Content Architecture
  - *Information in a block is coded by a single encoding scheme.*
    - Character Content Architecture (for characters)
    - Raster Graphics Content Architecture (for facsimile images)
    - Geometric Graphics Content (for computer graphics using CGM)
    - Audio Content Architecture (for audio)
    - ...

Processable/Formatted

- 以上説明してきた情報構造を利用する事により、多様なDocument 通信が実現できる。

- Formatted Document: Layout StructureとContent Architecture を送る方式。
  - *受信者は一切再利用ができない。*
- Prosessable Document: Logical StructureとContent Architecture を送る方式。
  - *受信者はLayouting Processを経ないとDocumentを見る事はできない。*
  - *再利用が可能。*
- Formatted Processable Document: 3つの構造全てを送る方式。
  - *上記2つを合わせ持った特徴を持つ。*

Processable/Formatted

- Using the those structures, various document communications can be realized:

- Formatted Document: method sending layout structure and content architecture.
  - *Receivers cannot re-use the documents.*
- Prosessable Document: method sending logical structure and content architecture.
  - *Receivers cannot see the document unless getting through the layout processing.*
  - *Re-using is possible.*
- Formatted Processable Document: method sending all three structured data.
  - *It has both characteristics of formatted and processable documents.*

ODIF

- Open/Office Document Interchange Format
- 上記述べてきた、Logical Structure, Layout Structure, Content Architecture表現を、合わせて符号化し交換するためのFormat。
- ASN.1で表現する、Binary Structured Data。

- ASN.1 (Abstract Syntax Notation No.1, 抽象構文記法1)
- ISO/IEC 8824, 8825に規定される。
- Structuring Data Representation
- 通信分野で非常に良く使われる複合Data記述手法。
  - *Dataの、集合・選択・属性等をきめ細かく指定できる万能なData記述言語*
- Data構造を決める文法（雛型）とその符号化を別々に指定できる。
  - *符号化方式: BER, DER, CER, PER*

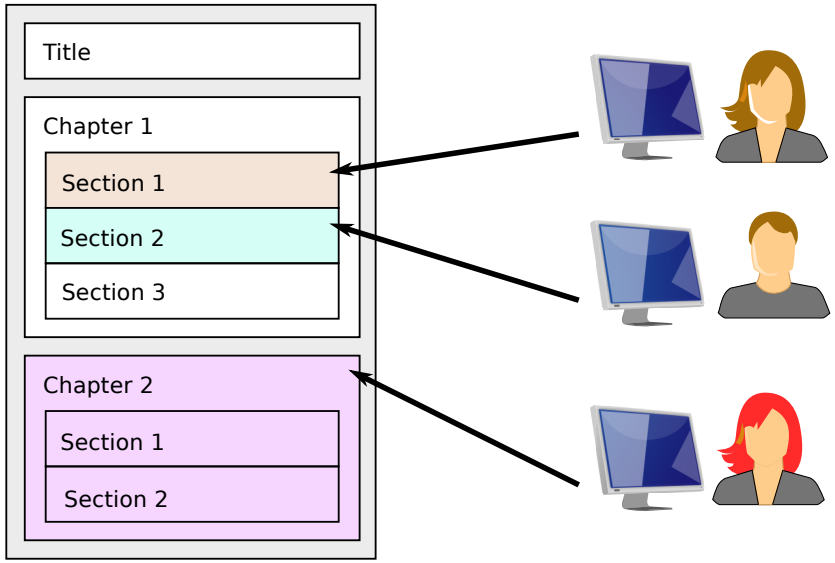
ODIF

- Open/Office Document Interchange Format
- A format used for encoding logical structure, layout structure and content architecture, in order to exchange document info.
- Binary structured data represented by ASN.1.

- ASN.1 (Abstract Syntax Notation No.1)
- Specified by ISO/IEC 8824 and 8825.
- Structuring Data Representation
- Composite date description method widely used in telecom applications.
  - *It is a general-purpose data description language with capabilities to represent "set", "selection of data", "attributes", and so on.*
- Grammar for structured data and its encoding rule are separately specified.
  - *Coding rules: BER, DER, CER, PER*

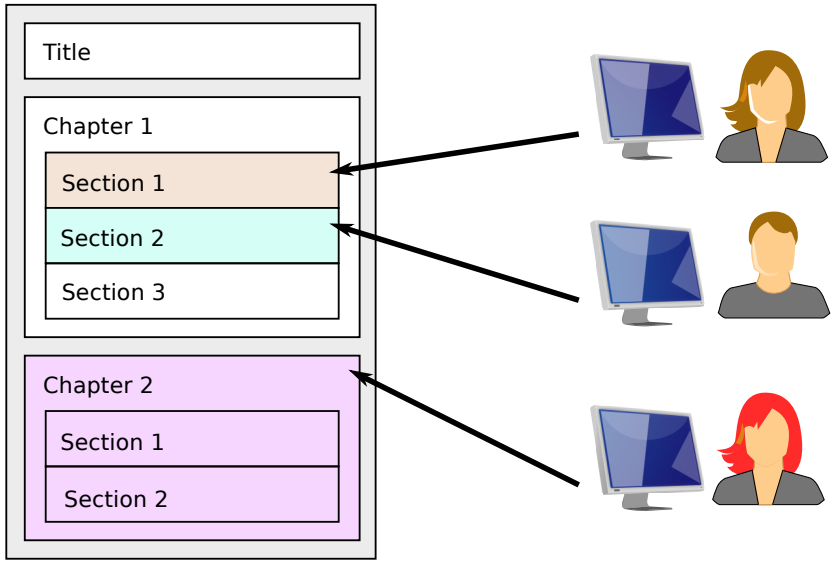
DTAM

- Document Transfer and Manipulation
- ODA文書をRemoteからEditingするためのProtocol
- 協調文書作成を念頭に考えられた。
- 今日の WebDAVや Delta-Vと同じ概念。
  - *WebDAV: RFC 4918 "HTTP Extensions for Web Distributed Authoring and Versioning (WebDAV)", June 2007*
  - *Delta-V: RFC 3253, "Versioning Extensions to WebDAV (Web Distributed Authoring and Versioning)", March 2002*

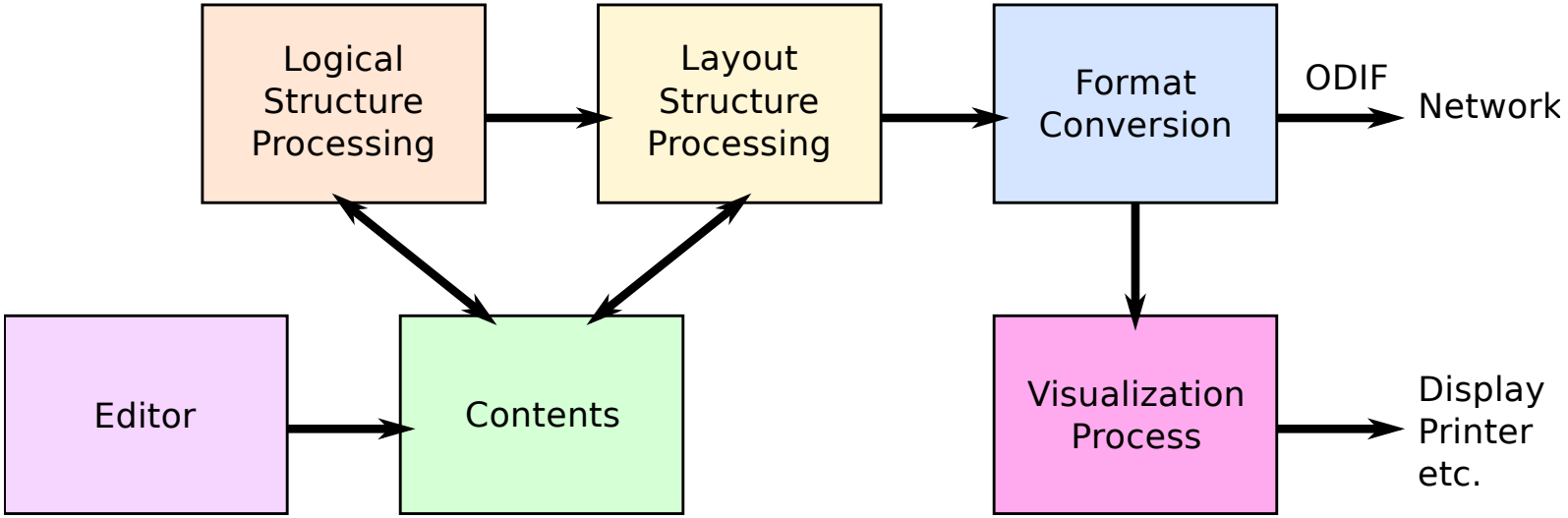


DTAM

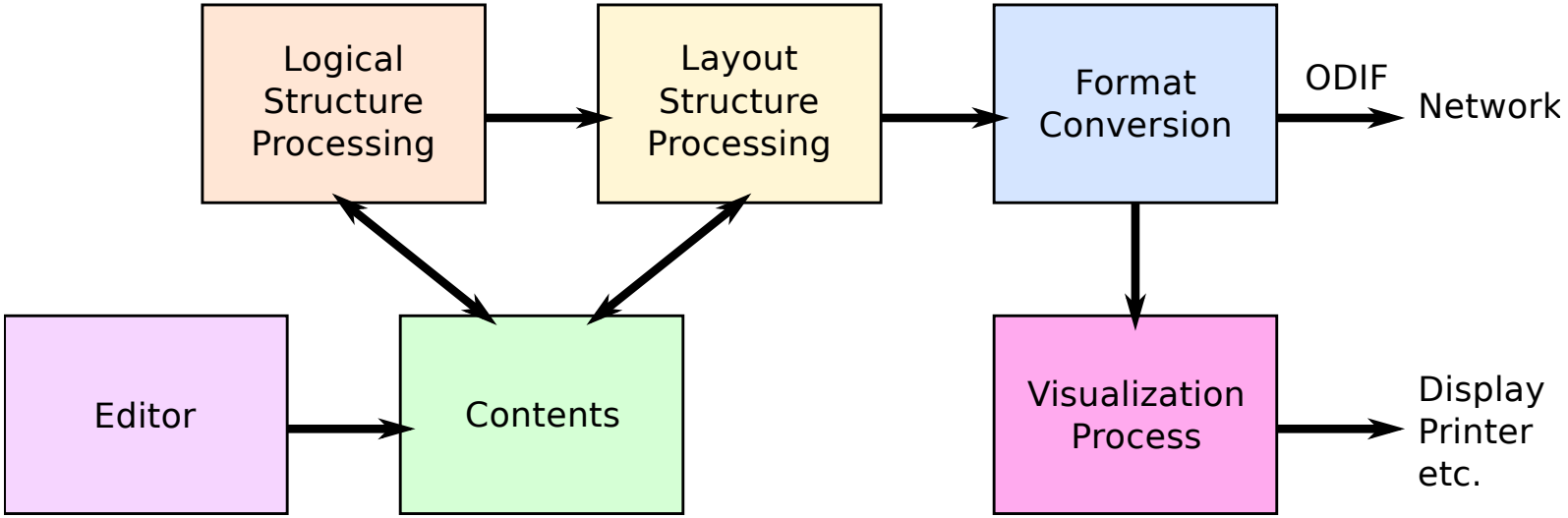
- Document Transfer and Manipulation
- A protocol used for remote editing of ODA documents.
- Enabling collaborative document processing and editing environment and system.
- Same as WebDAV and Delta-V as of today in terms of concept.
  - *WebDAV: RFC 4918 "HTTP Extensions for Web Distributed Authoring and Versioning (WebDAV)", June 2007*
  - *Delta-V: RFC 3253, "Versioning Extensions to WebDAV (Web Distributed Authoring and Versioning)", March 2002*



ODA文書処理Model



ODA Document Processing Model



ODAにおける情報交換Protocol

- MHS: Message Handling System (ITU-T)
  - *今日のMIME/SMTPと同等*
- MOTIS: Message Oriented Text Interchange System (ISO/IEC JTC1)
  - *今日のMIME/SMTPと同等*
- FTAM: File Transfer Access and Management
  - *今日のFTPと同等*

- 上記の内、いずれかを使用する。

Doc Exchange Protocols used in ODA

- MHS: Message Handling System (ITU-T)
  - *Same as MIME/SMTP of today*
- MOTIS: Message Oriented Text Interchange System (ISO/IEC JTC1)
  - *Same as MIME/SMTP of today*
- FTAM: File Transfer Access and Management
  - *Same as FTP of today*

- One of above is to be used.

なぜODAは受け入れられなかったのか

- 巨大な規格であった。
- 各国の思惑が渦巻き、Profile作成に失敗した。
- Word Processorの台頭
- OSIの挫折
- …

Why Hasn't ODA been Accepted?

- Too big specifications.
- Failed in making profiles due to too different requirements.
- Coming up of inexpensive word processors and PCs.
- OSI failure
- ...

まとめ

- Logical Structure/Layout Structureという重要な発見。
- 上記構造の基本的な考え方。
- Content Architectureを上記と全く独立に持っている事。
  - *Contentの符号化方式（表現方式）が変わってもLogical Structure/Layout Structureは変化しない。*
- 新しいApplicationに対する先駆的な考え方。
  - *Remote Document Editing*
  - *Visualization Processの独立化*
- Processableという考え方。

- ODAの考え方は、今日のMultimedia Processingの基本概念を全て押えている。

Wrap Up

- Very innovative discovery of "Logical Structure" and "Layout Structure".
- Basic concepts of above two structures.
- Independence of content architecture from above two structures.
  - *Identical logical/layout structures can be used if encoding (or representation) methods of contents are changed.*
- Innovative concept of digital document processing system and environment.
  - *Remote document editing.*
  - *Separation of visualization process.*
- An important concept of "Processable".

- ODA introduced all of the multimedia processing concepts and ideas currently we have.