Wikipedia

† †† † †,††† †

† 464-8601
†† 500-8288 1-38
††† 101-8430 2-1-2
E-mail: † {okuoka,ide,murase}@murase.m.is.nagoya-u.ac.jp, †† ttakahashi@gifu.shotoku.ac.jp, ††† ide@nii.ac.jp

Wikipedia

Wikipedia
86% 79%

Wikipedia

Automatic Indexing of a News Video Archive with Wikipedia Entries

Tomoki OKUOKA† Tomokazu TAKAHASHI†† Daisuke DEGUCHI†

Ichiro IDE†, ††† and Hiroshi MURASE†

† Nagoya University Graduate School of Information Science Furo-cho, Chikusa-ku, Nagoya, Aichi, 464-8601 Japan
†† Gifu Shotoku Gakuen University Faculty of Economics and Information 1-38, Nakauzura, Gifu-shi, Gifu, 500-8288 Japan
††† National Institute of Informatics 2-1-2 Hitotsubashi, Chiyoda-ku, Tokyo, 101-8430 Japan
E-mail: † {okuoka,ide,murase}@murase.m.is.nagoya-u.ac.jp, †† ttakahashi@gifu.shotoku.ac.jp, ††† ide@nii.ac.jp

Abstract Following the increase of video contents on TV or the Internet, efficient techniques to browse and search video data in an archive are needed. We propose a method to support browsing a news video archive with the help of Wikipedia because news videos are important as video contents. First, videos are automatically indexed by Wikipedia entries by means of evaluating the similarity of text information. Using those indices, we extract and present video groups via an interface, that correspond to each Wikipedia entry. Through experiments, news videos were accurately indexed by Wikipedia entries with a precision of 86% and a recall of 79%. In addition, we confirmed that we could obtain information which could not be obtained by closed captions, such as understanding the transition of topics or discovering the associations between videos which are broadcasted on distant days.

Keyword Wikipedia News Video Video Archive Automatic Indexing

1.

YouTube(*1)

NHK (*2)

Closed Caption ; CC

1 http://www.youtube.com/

2 http://www.nhk.or.jp/archives/

[1]

Wikipedia(*3) Wikipedia

Wikipedia

[1]

Wikipedia

2 1 Wikipedia

CC

Wikipedia

2

2 3

Wikipedia

4

5

[2]

2.

2.1. ニュース映像の閲覧に関する研究

[2][3]

1

Wu

2

[4]

[1]

1

Wikipedia

Rautiainen cluster-temporal browsing[5] Snoek MediaMill[6]

2.2. Wikipedia の利用に関する研究

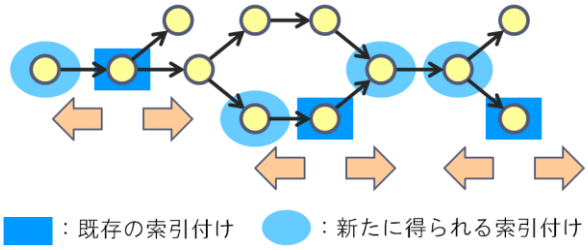
Wikipedia

Wikipedia

Wikipedia Web

[7]

3 http://ja.wikipedia.org/wiki/Wikipedia/



4

3.3.2.

[1]

4

Wikipedia

CC

3.4. 各エントリに関連する映像群の抽出

Wikipedia

[1]

Wikipedia

4.

4.1. 使用するデータ及び実験条件

CC NHK 7
2007 1 1 2008 6
30 CC CC

Wikipedia 2008 11 27

Wikipedia 1,053,561

1,645

4.2

4.3

4.4

4.2. 実験 1 ：対応付け精度

4.2.1.

Wikipedia CC

3

Wikipedia (＊5)

2007

CC Wikipedia

Wikipedia

CC

CC

2 3

1

2

3

3 1

1

	1	2	3
%	43.4	97.4	86.1
%	95.1	45.4	79.3

5

4.2.2.

2 3

1

CC

CC

1 3

2 Wikipedia

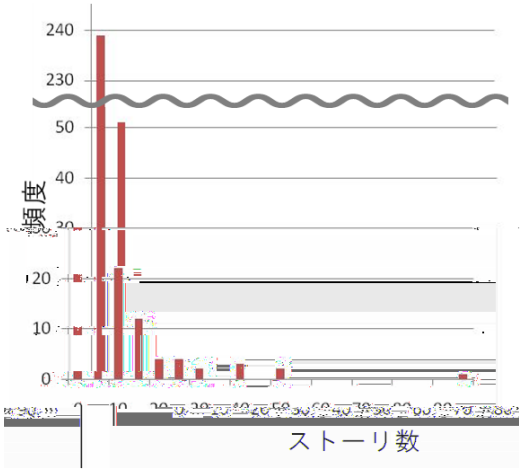
Wikipedia

11%

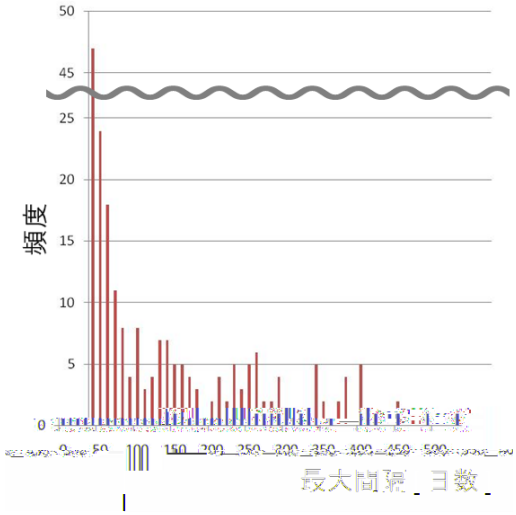
63%

Wikipedia

1 15%



5



6

4.3. 実験 2 :ストーリ数と最大間隔

4.3.1.

Wikipedia

Wikipedia

1,645 5 6

0 0

0 Wikipedia 1,305 79%

0 Wikipedia 1,421 86%

4.3.2.

Wikipedia

0 20

0

NHK 7 1

300

Wikipedia

6

NHK

498 2007 1 29

2008 6 10

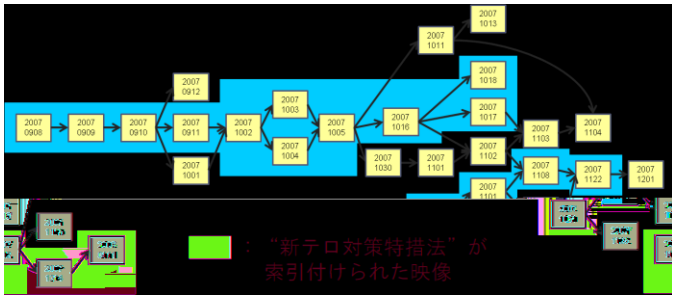
4.4. 映像群の抽出結果

4.4.1.

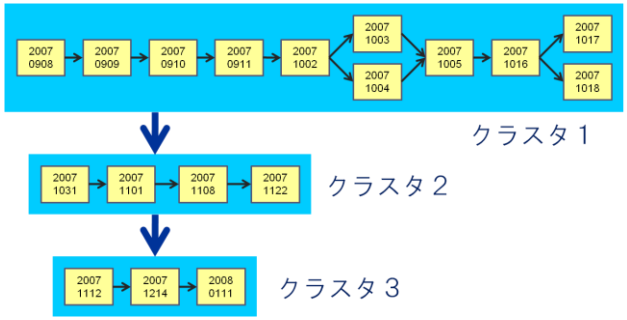
Wikipedia

Wikipedia 7

2007 9 8 3



7



8

8 Wikipedia

CC

4.4.2.

8

Wikipedia

1 2007 9 10

2 2007 11

2007

3 2007 12

Wikipedia

Wikipedia

5.

Wikipedia

86% 79%

498

Wikipedia

Wikipedia

Wikipedia

Wikipedia

Wikipedia

SlothLib

http://www.dl.kuis.kyoto-u.ac.jp/slothlib

[1]

Vol.23 No.5 pp.282 292 (Sep. 2008

[2] National Institute of Standards and Technologies: The year 2000 Topic Detection and Tracking (TDT2000) Task Definition and Evaluation Plan (2000), http://www.itl.nist.gov/iad/mig//tests/tdt/2000/

[3] P. Duygulu, J.-Y. Pan, and D. Forsyth Towards Auto-Documentary: Tracking the Evolution of News Stories”, Proc. 12th ACM Int. Conf. on Multimedia, pp.820 827 (Oct. 2004)

[4] X. Wu, C.-W. Ngo, and Q. Li Threading and Autodocumenting News Videos , IEEE Signal Processing Mag., Vol.23, No.2, pp.59 68 (Mar. 2006)

[5] M. Rautiainen, T. Ojala, and T. Seppanen “Cluster-Temporal Browsing of Large News Video Databases ”, Proc. 2004, IEEE Int. Conf. on Multimedia and Expo, Vol.2, pp.751 754 (June 2004)

[6] C. Snoek, M. Worring, J. van Gemert, J.-M. Geusebroek, D. Koelma, G. Nguyen, O. de Rooij, and F. Seinstra MediaMill: Exploring News Video Archives based on Learned Semantics , Proc. 13th ACM Int. Conf. on Multimedia, pp.225 226 (Nov. 2005)

[7] Wikipedia

Vol.47 No.10 pp.2917 2928 (Oct. 2006)

[8] Wikipedia

2008-NL-187 (pp.83 90) (Sep. 2008)