

HAIKU サーバのアクセスログの解析と運用管理

和田 武[†] 檀 裕也[‡] 墨岡 学[‡]
 愛媛大学[†] 松山大学[‡]

1. はじめに

郷土が誇る俳人正岡子規の HAIKU サーバ (SHIKI) を運用して 10 年が経過した。SHIKI では 2 つの英文俳句メーリングリスト (NOBO List, Shiki Haiku forum) とそのアーカイブ (Shiki old archives)、および HaikuSphere を運用している。SHIKI 運用開始以来、世界各国から活発に利用され、地域の特徴を活かした情報発信による地域社会への貢献および国際交流に貢献している。

HAIKU サーバを含む Web サーバのアクセスログを用いた研究は、単純にアクセスログを解析した報告や、多変量解析などの手法を用いてサイト利用者の閲覧行動の分析や特徴付けを行った研究など数多くの研究がなされている。

今回、HAIKU サーバのアクセスログを解析して、サイト利用者の閲覧行動の支援や、Web サイトの構造の見直しを図る目的など今後の運用管理に役立てるために SHIKI のアクセスログを解析した。

本稿では、調査期間中の (a) 月別・時間別アクセス状況、(b) トップレベルドメイン (以下、TLDs) 別アクセス状況について、前回の調査 (2002 年 12 月から 2005 年 6 月 2.5 年間) と比較しながら報告し、今後の運用計画について述べる。

2. 分析方法と分析結果

(1) 分析方法

分析データは、記録が残っている 2002 年 12 月から 2006 年 11 月までの期間のアクセスログを利用した。アクセスログは、アクセス日時、ページ、ディレクトリ、アクセスしたホスト名など膨大なデータから構成されている。

(2) 分析結果

分析期間中、2347.5 万件 (1 日あたり平均

1.62 万件) のアクセスがあり、トラフィック量は 137.8GB (同 95MB) であった。これは、愛媛大学各部署に設置している基幹サーバ中 4 番目に多いアクセス数 (しかも 1 台のサーバ) であり、非常によく利用されているサイトであることが確認された。

(a) 月別・時間別アクセス状況

分析期間中の月別アクセス状況について調査した。2002 年 12 月から 2006 年 6 月まで急激に増加しており、特に 2006 年に入ってからの増加が著しい。これは、google などの検索エンジンのロボットによるアクセスが影響しているものと思われ、今回新規に検索エンジンのロボットによるアクセスに絞って調査した。その結果、google が 50%、Yahoo!Slurp が 6%、MSN:3%、Ask:4% となり、google が群を抜いて多く、ロボットによるアクセスが多いことが検証された。

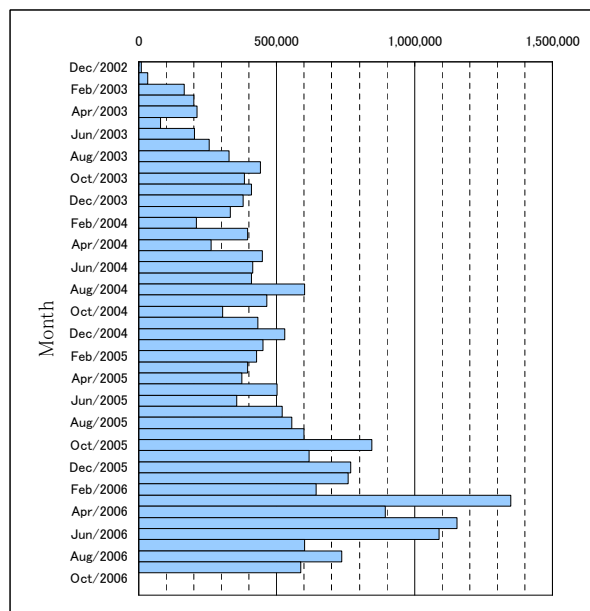


図 1. 月別アクセス状況

次に、時間別アクセス状況について調査した (図 2)。%Reqs はアクセス要求回数 (割合)、%Byte はトラフィック量 (割合) を示す。%Reqs の曲線に対する多項式近似曲線は、

$$y = -0.00006x^4 + 0.0027x^3 - 0.0355x^2 + 0.122x + 4.1534$$

Analysis of HAIKU server Access logs and management

[†] Takeshi Wada, Center for Information Technology, Ehime University,

[‡] Manabu Sumioka, Yuya Dan, Faculty of Business Administration, Matsuyama university

($R^2=0.8921$) となり、%Byte の曲線に対する多項式近似曲線は、

$$y = -0.00004x^4 - 0.0025x^3 + 0.0622x^2 - 0.5841x + 5.5927$$

($R^2=0.7331$) となった。これらの傾向は前回調査の時と同様で、時間帯による差は認められなかった。これは、全世界からアクセスされているためである。ちなみに、同期間中に国内に限って調査したところ、9時から24時の時間帯のアクセスが多く、2時から8時まではアクセスが少なく、国内の利用者の行動パターンが実証された。

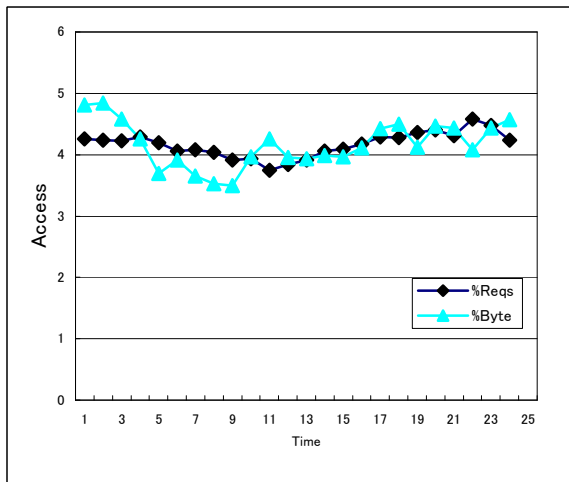


図2. 時間別アクセス状況

(b) TLDs 別アクセス状況

分析期間中に Web サーバにアクセスしたトップレベルドメイン(TLDs)を調査した。アクセスがあった167個のTLDsの中でリクエスト回数が2000回以上のドメインが45個あり、そのうちの上位20を表1に示す。Requestsは要求があった回数を、DomainはアクセスがあったTLDsを示している。表1の通り、comドメインが74.6%と圧倒的に多く、net, jp, nl, no, orgと続いている。トップ20をみただけでも世界各国から利用されていることがわかる。

これらの分析の結果、サーバの負荷分散やセキュリティ対策等は当面現状維持とし、さらに検討を重ねていくこととする。

3. おわりに

本稿では、HAIKUサーバSHIKIのサイト構造の見直しや、サイトを訪れた利用者の行動支援等の運用管理を行う目的で、アクセスログの解析を行った結果について述べた。このサーバは、1994年に開始されたShiki Internet Haiku Salonを原点として愛用されて現在に至っている。世界

160を超える国からアクセスがあり、1日平均1.62万件、95MBを超えるアクセスがあり、活発に利用されていることがわかった。アクセスログ分析の結果、サイトの運用は当面現状維持の方向で継続することとした。

今後は、オンライン俳句を世界各国で共有し楽しむためのHAIKU Wiki (Webブラウザからオンライン俳句ページの編集等が簡単に行え、柔軟性が高く操作が簡単なコンテンツ管理システムを利用した世界の短詩型文学の専用サイト) や、HAIKU SNS (世界各国の俳人が互いに幅広いコミュニケーションを取り合うことを目的とした招待型のWebサイト)のサイト運用を目指したい。

表1. TLDs別集計 (Top20)

No	%Reqs	%Byte	Requests	Domain
1	74.64	41.01	17521115	com
2	3.97	11.31	931853	Net
3	3.24	25.44	761464	jp
4	0.92	3.03	215627	nl
5	0.38	0.25	89293	no
6	0.2	0.2	46950	Org
7	0.2	0.41	46391	Edu
8	0.18	0.23	41267	uk
9	0.12	0.22	27337	ca
10	0.12	0.3	27287	it
11	0.08	0.14	18748	au
12	0.07	0.26	16955	de
13	0.06	0.1	14352	tw
14	0.06	0.07	13466	Edu
15	0.06	0.19	12922	fr
16	0.05	0.13	11643	be
17	0.05	0.08	10823	pl
18	0.04	0.04	10461	fi
19	0.04	0.09	10421	us
20	0.04	0.09	8581	ro

本研究にご協力頂いた、Shiki チームのメンバーである、井上博巳氏、田中喜美代氏他の皆様に感謝します。

4. 参考文献

- (1) 和田武、墨岡学、俳句サーバの運用管理とキーワードの定量的分析、大学情報システム環境研究, Vol. 9、pp71-74, 2006. 3.
- (2) 和田武、墨岡学、インターネット俳句サーバSHIKIの運用と効果、大学情報システム環境研究, VOL. 7、pp43-46, 2004. 3