

地方創生に向けた観光施策の立案・実施におけるクラウド AI の活用可能性

IA-005：クラウド AI の活用術：大熊 正浩

Availability of Cloud AI in planning and conducting tourism measure for regional revitalization

How to utilize cloud AI : Masahiro Okuma

政府は地方創生を推進しており、その中でも観光振興は重要な施策であるが、観光振興には観光動向の把握が必要である。しかし、従来から観光動向の把握に用いられてきた統計的手法にはいくつかの問題がある。本稿ではクラウド事業者が提供する AI のサービスを活用し、画像解析と自然言語処理を組み合わせた分析を行うことによって、観光動向の把握に関する問題の解決策を検討し、一定の効果があることを確認した。

The government promotes regional revitalization, particularly, although tourism promotion is an important measure, it is necessary to grasp tourism trends for tourism promotion.

However, the statistical methods conventionally used for grasping tourism trends have some problems.

In this paper, we will consider the solution for problems to grasp tourism trends by performing analysis combining image analysis and natural language processing using AI services provided by cloud companies and it was confirmed that the solution had certain effects.

Key Words & Phrases : 観光, クラウド AI, 画像解析, 自然言語処理, クラスタリング
tourism, cloud AI, image analysis, NLP, clustering

1. はじめに

近年、日本は人口減少・超高齢化という問題に直面している。この傾向は地方部でより顕著なため、政府は平成 26 年に、地方自治体と一体となって地方創生を推進する「まち・ひと・しごと創生総合戦略」[1]を閣議決定し、地域活性化を目指している。中でも観光振興は東京オリンピックを前にした外国人観光客の増加と相まって、重要な政策の一つとされている。政策会議の一つ「明日の日本を支える観光ビジョン構想会議」[2]では、「我が国の豊富で多様な観光資源を、誇りを持って磨き上げ、その価値を日本人にも外国人にも分かりやすく伝えていくこと」が課題として挙げられている。

これらの実現には観光動向の把握が不可欠である。多くの場合、統計を活用した手法が取られるが、時間とコストがかかる等、いくつかの問題がある。

このような施策に係る問題の解決手段として、近年は AI が注目を集めている。しかし、AI の活用には高い専門性が必要で、人材は不足している。ところが、最近になり、クラウド事業者から手軽に利用可能なクラウド AI が提供されるようになった。そこで筆者は、クラウド AI を活用し、観光施策立案、実施における観光動

向の把握について、問題解決の手法を検討する。

なお、本稿において、クラウド AI とはクラウド事業者の提供する AI サービスと定義する。

2. 観光施策に係る課題

2.1 統計の活用と問題

各地域の観光施策に統計が活用されている。現在、観光庁でも観光施策に役立てるため、各都道府県の観光入込客統計[3]（以下、「観光庁統計」と記載）の集計を約 1 年かけて行い、現状の把握に努めている。

しかし、統計を活用することには、二つの問題があると考えられる。

一つ目は、統計には時間とコストがかかる問題である。実際に観光庁統計では、3 年前の統計値においてさえ、一部集計中のものがある。また、「平成 31 年度 観光庁関係予算概要」[4]によれば、統計の整備に 6 億円以上の予算が計上されている。

二つ目は、統計では定性的かつ詳細な情報が見えなくなってしまう問題である。一般的に統計では、定義された属性項目に対し情報を数字として集約するため、定性的で具体的な情報は表現されなくなってしまう。

提出日：2019 年 7 月 31 日

IA005：クラウド AI の活用術(IA005: How to utilize cloud AI)

2.2 問題解決可能な手法の検討

上記を踏まえれば、観光動向の把握のために、タイムリーかつ低コストにデータを分類、集計できる手法、あるいは、定性的で解像度の高い情報を取得可能な手法が必要と考える。

そこで、本稿では、そのような手法を実現する手段として、SNS から得た旅行画像を、クラウド AI により解析することで有用な知見を得る手法を検討し、検証した結果について論じる。

DAC 社の調査「旅好き会員に聞く「旅先選びの SNS 活用」実態調査」[5]では、観光で SNS を活用する人は半数以上との結果があり、SNS に投稿された旅行画像は観光施策立案、実施のための有効な材料と考える。

以降、3 章では、各地域の観光地としての傾向把握を目的に、画像解析が観光庁統計を代替し得るか検証した結果について述べる。

4 章では、統計としては表現されない情報の活用手法として、画像解析結果から埋もれていた観光資源の発掘が可能か検証した結果について述べる。

最後に 5 章で本稿をまとめる。

3. 画像解析を用いた観光に関する傾向の把握

本章では、観光庁統計により示された観光地点ごとの入込客数から把握できる傾向を、クラウド AI を活用した画像解析から把握することが可能か検証する。これが可能であれば、統計を一部代替することができるため、統計を活用した観光動向の把握にかかる時間とコストを削減できる。

3.1 仮説

画像解析を目的としたクラウド AI は、画像に写っている物体を識別しその物体名を出力する機能を持つ。

都道府県ごとの旅行画像を画像解析 AI で解析し、出力された物体名からその画像が観光庁統計における観光地点分類（「観光入込客統計に関する共通基準」[6] P.10-P.11）のどれに当てはまるかを判定できれば、観光地点分類に従った画像の分類が可能である。分類結果から、観光庁統計と同様の傾向を得られることを期待し、これを仮説として設定した。

3.2 手法

画像がどの観光地点分類に該当するかの判定は、クラウド AI による画像解析から得られた物

体名と、観光地点分類の小分類（以下、「観光地点小分類」と記載）との類似度を測定することで実施した。

物体名および観光地点小分類は共に単語の集合であるため、自然言語処理の手法を用い、単語の集合同士の類似度を測定した。

本検証では、平成 29 年度の「都道府県別観光地点・行祭事・イベント別観光入込客数（延べ）」[7]の集計値と画像分類結果を比較した。対象とする都道府県は、同年度の入込客数上位 3 県である「東京都」、「福岡県」、「神奈川県」とした。

3.2.1 選定したクラウド AI

大手クラウド 4 社（Amazon Web Service(AWS)、Microsoft、IBM、Google）の画像解析 AI を比較検討し、AWS の「Amazon Rekognition」を選定した。自然言語処理では、低頻度語が OOV(Out Of Vocabulary)となり対処が必要となる課題があるが、旅行画像 100 枚に対し 4 社の AI で解析した結果を比較したところ、Amazon Rekognition が物体名の抽象度が適切で、低頻度語が少ないと判断したためである。

3.2.2 画像の具体的分類手順

- (1) 画像共有 SNS「Flickr」から各都道府県の画像（平成 29 年撮影）を 400 枚取得。母団体の傾向を十分反映した標本数は 385 枚（標本誤差 5%、信頼区間 95%）だが、解析エラーの発生を考慮し余裕を持たせるため 400 枚とした。なお、Flickr を選んだのは API が充実し画像取得が容易なためである。
- (2) Amazon Rekognition にて画像解析。
- (3) 解析結果から物体名を抽出。なお、抽出したのは解析結果の確信度が 80%以上の物体名のみ。
- (4) 抽出した物体名の集合と、観光地点小分類との類似度を測定。測定方法は下記のとおり。
 - a. Word2Vec を用い、1 枚の写真の解析結果に含まれる物体名ごとの特徴ベクトルの平均 (A) と、観光地点小分類の単語ごとの特徴ベクトルの平均 (B) を算出。物体名は英語で出力されるため、観光地点小分類も英語化した（表 1 参照）。また、特徴ベクトルを算出するために必要なモデルとして、fastText の学習済みモデル[8]を使用した。
 - b. 上記 A と B のコサイン類似度を算出。
 - c. 最も類似度の高かった観光地点分類を当該画像の属する分類とする。なお、どの

JGS 研究 2019 プロジェクト・チーム論文

観光地点分類とも類似度が 50%未満だった画像は「その他」に分類。

(5) OOV が発生した場合、手作業にて物体名を類義語に修正し、再度類似度を測定。

表 1. 観光地点小分類（英語化）

観光地点分類	小分類
①自然	Mountain, Plateau, Wetland, Wilderness, Lake, River, Canyon, Waterfall, Coast, Sand, Cape, Underwater, Island
②歴史・文化	Historic, Tombs, Castle, Ruins, Shrines, Temples, Gardens, Museum, Art, Zoo, Botanical, Aquarium, Brewery
③温泉・健康	Spa, Hot-spring, Health, Bath
④スポーツ・レクリエーション	Sports, Recreation, Golf, Ski, Campground, Fishing, Beach, Marina, Park, Leisure, Amusement
⑤都市型観光	Commercial, Mall, Shopping, District, Street, Food, Gourmet
⑥その他	-

3.3 結果

まず、各都県における観光地点分類ごとの画像分類結果を表 2 に示す。

表 2. 観光地点分類ごとの画像分類結果

単位：枚

	①	②	③	④	⑤	⑥
東京都	65	31	1	36	218	43
福岡県	100	43	1	32	174	46
神奈川県	168	26	0	28	126	40

※合計が 400 枚にならないのは解析エラーとなった画像が存在したため

次に、観光庁統計と画像分類結果を比較する。両者の差異を視覚的に示すため、両者を比較したレーダーチャートを図 1 に示す。チャートの尺度は観光地点分類ごとの構成比である。

3.4 評価

図 1 のとおり、東京都、福岡県および神奈川県のいずれにおいても、観光庁統計と画像分類結果との間で傾向が一致するのは一つないしは二つである。観光庁統計を正とすれば、本手法

により観光地としての傾向を精緻に把握可能とはいえず、観光庁統計を代替できるものではない。

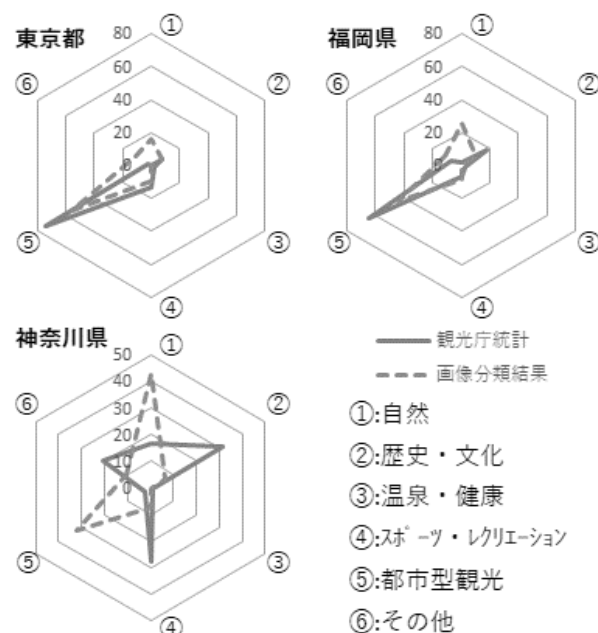


図 1. 観光庁統計と分類結果比較

一方で、定義となる単語群により画像を分類することには成功した。

図 2 は観光庁統計との相違が最も大きかった神奈川県の画像分類結果を Graphviz というツールを使って可視化したものである。例として件数上位の「自然」および「都市型観光」の画像を示すが、高い精度で分類できた。

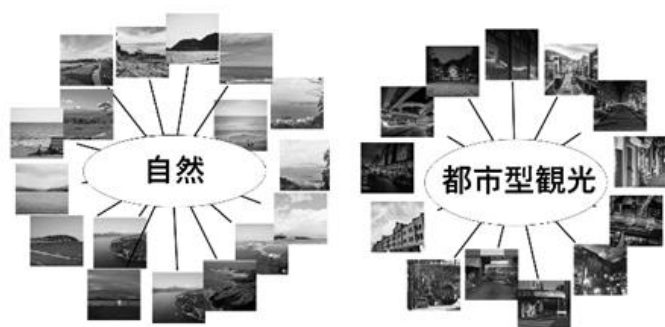


図 2. 神奈川県の自然および都市型観光の画像

観光庁統計と画像分類結果との相違の理由として下記が考えられる。

- (1) インプットとした画像の偏り
プライバシーの問題から SNS には温泉の画像をあげづらい等、偏りがあるのは否めない。
 - (2) 画像解析結果と観光地点小分類とのマッチングが高難度
例えば公園の木々の画像の場合、分類の定義
- JGS 研究 2019 プロジェクト・チーム論文

から「スポーツ・レクリエーション」に分類されるべきだが、木々だけを見れば「自然」に分類される可能性がある。正しい分類にはどこの木々かを考慮すべきだが、画像だけでは十分な情報を得られず、難易度が高い。

4. 画像集合のクラスタリングによる観光資源の発掘

本章では、クラウド AI の画像解析結果を活用したクラスタリングにより、統計では表現されなかった情報を活用し、埋もれていた観光資源の発掘が可能か検証する。

4.1 仮説

特定の都道府県の旅行画像を画像解析 AI で解析し、その結果を用いて画像群をクラスタリングする。得られたグループの特徴を見出し、統計では把握できない埋もれた観光資源が発掘されることを仮説として設定する。

4.2 手法

画像に対し、そこに含まれる物体名およびそれらの確信度を用いてベクトルを定義し、ベクトルのコサイン類似度によってクラスタリングを行う。

画像集合のクラスタリング手法として、画像そのものの類似度を利用する方法がある。しかし、直接画像間の類似度を計算するためには、予め大量の画像を収集した上で、人手により教師データを作成し、Deep Learning などの手法を使って類似判定器を構築する必要がある。この手法は高度な AI の技術、知見が必要であり、本稿の趣旨とは異なるため、採用しない。

旅行写真に対するクラウド AI の物体認識結果を用いた先行研究[9]では、各画像に含まれる物体名を確信度でベクトル化することで物体名をクラスタリングする手法をとっている。

本稿ではこれを応用し、物体名ではなく画像を確信度でベクトル化し、画像をクラスタリングする手法を考案、採用した。

検証対象として鳥取県を選定した。鳥取県は観光庁統計において、入込客数が下位に位置しており、施策の立案手法に改善の余地があると考えたからである。

鳥取県は、平成 27 年に「鳥取県観光振興指針（計画期間：平成 27 年度～30 年度）」[10]

（以下、「観光振興指針」と記載）を発行した。これは、各種統計資料等を参考に、鳥取県が観光振興施策を策定し、指針としてまとめたものである。

観光振興指針にない観光資源を発掘できれば、仮説を立証できたといえる。

4.2.1 選定したクラウド AI

大手クラウド事業者のサービスを比較検討し、Google の「Google Cloud Vision API」を選定した。

見えなくなった情報を掘り起こし活用するためには、解析結果に含まれる物体名としてより具体的な情報が抽出されることが望ましい。当該サービスを選定したのは、Google の画像検索を活用した具体的な物体名が検出可能なウェブ検索オプションが用意されており、4 社の中で最も具体的な情報を得られたためである。

4.2.2 クラスタリング手順

以下、解析対象の画像枚数を n 、各画像を $P_i (1 \leq i \leq n)$ 、各画像から抽出された物体名の総数（異なる画像に出現する同一物体名を一つと数える）を m とする。

- (1) m 個の物体名に対し、各物体名をベクトル要素 $W_1 \sim W_m$ として定義。
- (2) i 番目の画像 P_i に対し、物体名 $W_j (1 \leq j \leq m)$ の確信度 c_{ij} とするとき、 c_{ij} を用いて以下の特徴ベクトルを定義。

$$\vec{P}_i = (c_{i1}, c_{i2}, \dots, c_{im})$$

なお、 W_j に対する確信度が P_i において未定義の場合、もしくは閾値 α （データとしての信頼度を考慮し 0.3 とした）を下回る場合は $c_{ij}=0$ とする。

- (3) 上記により定義したベクトルのコサイン類似度により、任意の 2 つの画像 $P_x, P_y (1 \leq x, y \leq n)$ の類似度を算出。
- (4) すべての画像をノードとした無向グラフを作成。各ノードは閾値 β （視認性向上のためエッジ数が 200 となる β を設定した）を超える類似度を持つ他のノードとエッジで結びグラフを表現。これによりノード（画像）がグループ化される。

4.2.3 特徴の見出し方

あるグループに対し、そこに属する各ノード（画像）の解析結果に含まれる物体名のうち、

JGS 研究 2019 プロジェクト・チーム論文

出現回数の多いものを、当該グループを特徴づける物体名とする。

4.3 結果

3 章と同様に Flickr から 500 枚の観光画像を取得し、先に挙げた手法でクラスタリングを行った。図 3 はクラスタリング結果を Graphviz により可視化したもので、グループ A～H が得られていることがわかる。

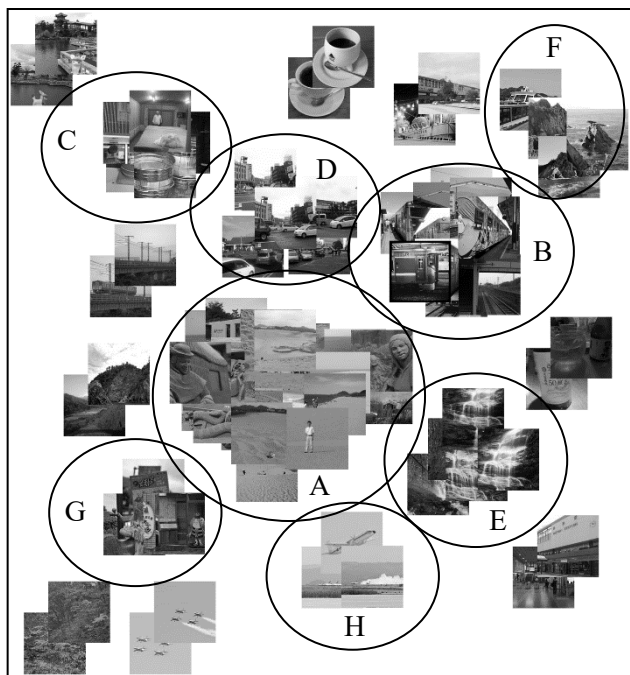


図 3. 画像集合のクラスタリング結果

なお、他のいずれのノードとも接続されないノードについてはグラフから除外している。

表 3 は各グループを特徴づける物体名を示したものである。

表 3. 各グループを特徴づける物体名

グループ	特徴づける物体名
A	Tottori Sand Dunes,Uradome Coast,秋吉台
B	Rail transport, Train, Railroad car
C	Chiyomusubi Okasora ,Honten, Sake, Soju
D	Luxury vehicle,Compact car,Mid-size car
E	Waterfall,Nature reserve,Water resources
F	Uradome Coast,Ginzan, San'in region
G	GeGeGe no Kitaro, Mizuki Shigeru Road,Yokai Shrine
H	Narrow-body aircraft,Aviation, Aircraft

4.4 評価

観光振興指針には、下記の観光資源を積極的に発信していくことが述べられている。

- ①鳥取砂丘、②山陰海岸ジオパーク、③大山、④かに・食、⑤まんが

A～H の各グループと観光資源①～⑤を、「特徴づける物体名」をキーとして対応づけたものが表 4 である。

表 4. 各グループと観光資源との関係

グループ	観光資源	説明
A	①	鳥取砂丘を示す画像群
B		旅行画像に該当しない鉄道画像
C		鳥取県酒造組合では利き酒大会の開催など、観光資源としての PR 活動に力を注いでいる
D		旅行画像に該当しない車両画像
E	③	「滝」を表す一般的な用語が物体名として選出されているが、画像は鳥取県東伯郡琴浦町の観光名所「大山滝」
F	②	浦富海岸は山陰海岸ジオパークに含まれる海岸で、鳥取県の観光名所
G	⑤	水木しげるロードをはじめとする漫画に由来する観光名所
H		鳥取県の航空自衛隊美保基地の航空ショーの様子

グループ C および H は、観光振興指針にはないものの、旅行者の興味を引く観光資源と考えられる。これは、観光振興指針では見いだせなかった埋もれた観光資源の発掘といえる。

また、表 4 からは、観光振興指針と一致する観光資源が四つあることがわかった。これは、観光振興指針に示した観光資源に多くの人が訪れていることを示しており、その効果が確認できたといえる。

更に、観光資源④についてはいずれのグループにも該当しないことがわかった。これは、観光振興指針の狙いどおりには人が集まっていないことを示し、効果の薄い施策の炙り出しができたといえる。

5. まとめ

5.1 検証結果の総評

3章では、観光庁統計から得られる各都道府県の傾向を、クラウドAIを活用した旅行画像の解析から把握可能かを検証した。結果として、画像解析から得られた傾向は、統計から把握した傾向と一致度が高いとはいえず、統計の一部を代替するには至らなかった。

しかし、画像解析と自然言語処理の組み合わせによる旅行画像の分類に成功した。分類結果から観光動向に係る何らかの知見を得ることは可能と考える。

4章では、鳥取県を対象に、クラウドAIを活用した旅行画像の解析から、埋もれた観光資源が発掘可能かを検証した。結果として、県の観光施策では見いだせなかった観光資源を二つ発見できた。更に、施策の効果確認や、旅行者に浸透していない施策の炙り出しにも成功した。

総評として、クラウドAIの活用が観光動向の把握に対し一定の有用性を持ち、観光施策立案、実施に役立つ可能性があることを確認できた。

5.2 クラウドAIのコストメリット

統計の一部を代替するには至らず、観光動向の把握にかかるコスト削減が可能とはいえないものの、クラウドAIにはコストメリットがある。

クラウドAIは実装が不要で、安い利用料で使用可能である。Amazon Rekognition は画像 1000 枚あたり 1.3USD、Google Cloud Vision API は 1000 枚以下なら無料である。

なお、検証の仕組み全体を構築するのに要した工数はおおよそ 1 人月である。画像取得、自然言語処理、およびクラスタリングについて、Python、Node.js、SQL を使い手組みで実装したが、簡易な仕組みのため工数は小さい。分類やクラスタリングのインプットに画像そのものではなくクラウドAIの解析結果（単語や数値）を使用できるため、簡易な仕組みで実現できた。

5.3 今後の課題

3章でも述べたとおり、SNS から取得可能な画像には偏りがある。偏りのない画像データの収集方法の確立は、本稿にて検証した手法を実用化するにあたっての課題であると考ええる。

また、統計の代替に向けては、画像解析結果を他の情報と組み合わせたり、より高度な自然言語処理を導入したりといった新たな手法の検討が必要となる。これらについては今後の研究

に期待したい。

謝辞

本稿の執筆にあたっては、日本アイ・ビー・エム株式会社 鈴木博和様より多くの助言を頂きました。改めて深謝致します。

参考文献

- [1] まち・ひと・しごと創生総合戦略について、
https://www.kantei.go.jp/jp/singi/sousei/meeting/keizai_kondan/siryo5.pdf, 2019.7.20
- [2] 「明日の日本を支える観光ビジョン」－世界が訪れたい日本へー 概要（案）,
https://www.kantei.go.jp/jp/singi/kanko_vision/dai2/siryou1.pdf, 2019.7.20
- [3] 共通基準による観光入込客統計,
<https://www.mlit.go.jp/kankocho/siryou/toukei/irikomi.html>, 2019.6.14
- [4] 平成 31 年度 観光庁関係予算概要,
<https://www.mlit.go.jp/common/001270653.pdf>, 2019.7.27
- [5] 旅好き会員に聞く「旅先選びの SNS 活用」実態調査,
<https://www.dac-group.co.jp/wp/wp-content/uploads/2018/06/a987e44ad9512d938ec805e9f1fd165.pdf>, 2019.7.6
- [6] 観光入込客統計に関する共通基準,
<https://www.mlit.go.jp/common/000995211.pdf>, 2019.6.14
- [7] 全国観光入込客統計に関する共通基準 集計表 調査時期：平成 29 年 「5.都道府県別観光地点・行祭事・イベント別観光入込客数（延べ）」,
<https://www.mlit.go.jp/common/001296060.xlsx>, 2019.6.14
- [8] fastText, English word vectors,
<https://fasttext.cc/docs/en/crawl-vectors.html>, 2019.6.14
- [9] 北村理紗, 伊藤貴之, 個人の旅行写真の一般物体認識に基づく観光地推薦, 映像情報メディア学会技術報告 = ITE technical report 42(12), 185-188, 2018.03 月
- [10] 鳥取県観光振興指針,
https://www.pref.tottori.lg.jp/secure/965427/shishin_201503.pdf, 2019.6.14

添付 1 : メンバーリスト (会社名アイウエオ順)

添付 2 : Member list

(alphabetical order of company name)

J G S 研究 2019 プロジェクト・チーム論文

添付1：メンバーリスト（会社名アイウエオ順）

会社名	部門・役職	名前	L/S
(株) インテック	D Xソリューションセンター	増田 浩士	
N D I ソリューションズ (株)	サービス事業本部 カスタマー I T 推進部	金澤 優太	
カリツー (株)	情報システム部	高木 浩行	
S O M P O システムズ (株)	ソリューション第 1 グループ	山崎 義則	(S)
第一生命情報システム (株)	基盤システム第二部 基盤運用第一グループ	中村 亮一	
パーソルテクノロジースタッフ (株)	技術本部 I T 第 1 技術部 システム開発 G	安田 遼平	
みずほ情報総研 (株)	技術企画部	大熊 正浩	(L)
三菱総研 D C S (株)	基盤サービス部	林田 祐樹	(S)

チームアドバイザー

会社名	部門・役職	名前
日本アイ・ビー・エム (株)	グローバルビジネスサービス事業本部 コグニティブプロセス変革	鈴木 博和

添付 2 : Member list(alphabetical order of company name)

Company	Name Given name Surname	L/S
INTEC Inc.	Hiroshi Masuda	
KARITSU CO.,LTD.	Hiroyuki Takagi	
Mitsubishi Research Institute DCS Co., Ltd.	Yuki Hayashida	(S)
Mizuho Information & Research Institute, Inc.	Masahiro Okuma	(L)
NDI SOLUTIONS LTD.	Yuta Kanazawa	
PERSOL TECHNOLOGY STAFF CO., LTD.	Ryohei Yasuda	
Sompo Systems Inc.	Yoshinori Yamazaki	(S)
The Dai-ichi Life Information Systems Co.,Ltd.	Ryoichi Nakamura	

Team adviser

IBM Japan, Ltd.	Hirokazu Suzuki
-----------------	-----------------