

引用情報に基づく基本特許抽出

小川 知也 渡部 勇

{ogawa.tomoya, watanabe.isamu}@jp.fujitsu.com

富士通研究所

〒 211-8588 川崎市中原区上小田中 4-1-1

本稿では、引用情報に基づく基本特許などの重要特許抽出実験について報告する。特許からの引例抽出、引例に基づく重要特許抽出、引用情報の可視化について述べる。国内特許約 530 万件から発明者引例、審査官引例の抽出を行い、それぞれの性質に違いがあることを明らかにした。引例の被引用数に基づく重要特許抽出実験を行った結果、特許庁サイトの報告書における重要特許の多くを抽出することが出来た。重要特許間の時系列的関係を表現する、引用情報に基づく引用フローによる可視化手法を提案し、実際のデータで有効性を確認した。

Finding Basic Patents using Patent Citations

Tomoya OGAWA Isamu WATANABE

Fujitsu Laboratories Ltd.

4-1-1, Kamikodanaka, Nakahara-ku, Kawasaki, Kanagawa, 211-8588 Japan

In this paper, we describe some experiments for finding basic patents using patent citations. We explain patent citation extraction from patents, the method for finding basic patents using patent citations, and the visualization of citations.

By extracting both inventor's and examiner's citations from 5.3 million Japanese patent documents, we have shown that there are differences between their characteristics.

We have been able to find many of the important patents in the reports at Japan Patent Office, by evaluating patents according to the citation frequency.

We propose a novel visualizing method based on citations, and show its effectiveness.

1 はじめに

論文、特許、技術報告書など電子化された技術文書が年々増大し、ネットを通じて入手可能となってきた。それらの技術文書を利用して技術のサーベイや特許の動向調査を行う場合、重要な技術とそれらの間の関係を知ることが必要となる。

技術文書の中でも特許は、侵害訴訟やライセンス供与など組織にとり大きな経済的影響を及ぼす可能性のある文書であり、基本特許などの重要特許抽出のような特許を対象にした分析技術の開発が重要性を増している。

重要特許抽出に関しては論文と同様に、重要な技術に関する文書は被引用数が多い傾向にあると言われる。折しも平成14年9月に設けられた「先行技術文献情報開示制度」により、先行技術文献情報を特許出願明細書に記載する必要があるとされ、今後は特許における引用情報である引例が増えることが予想される。

本稿では、国内特許を対象にした引例抽出、抽出した引例に基づく重要特許抽出、および重要特許間の関係の可視化に関する実験について報告する。

2 特許からの引例抽出

本節では、特許からの引例抽出について説明する。

国内特許では先行技術との違いを明確にし特許性を主張するために、発明者により出願時に先行技術文献が引例として明細書に記述される。また登録時には審査官により関連のある特許が付与される。本稿では前者を発明者引例、後者を審査官引例と呼ぶ。

2.1 発明者引例の抽出

発明者引例は、特許明細書に次のように記述される。

【0024】また、特開平5-305586号公報には、脚式移動ロボットの傾斜姿勢をZMP目標位置によって制御する点について開示している。

発明者引例はこの例「特開平5-305586」のように独特の表現であるため、文字列のパターンマッチにより抽出を行う。ただし次の2つの問題点がある。

- 列挙型の記述
- 特許番号での記述

複数の特許を列挙する時、次の例の「同60-140341」のように一部が省略される場合がある。

【0031】一方、有機シロキサン化合物は、写真用印刷紙、写真用フィルムの表面特性、例えば、べたつき感、くっつき耐性、耐傷性、すべり性等の向上を目的としてハロゲン化銀写真感光材料に従来より用いられてきた。特開昭50-117414号、同60-140341号、同60-191240号、同62-203152号、同62-203155号、特開平2-293843号等にそれらの適用について詳細に記されている。
...

このような列挙型の記述に対しては、省略されていない直前の番号を用いて省略された情報を適宜補充するようにした。

「特許3148830」のように特許番号で記述される場合、「米国特許3148830」の一部なのかそうではないのかの判断が必要となる。

特許番号での記述に対しては、直前が国名かどうかなどを判断することで多くの場合適切な結果が得られることが予備調査より分かったため、それにより対処した。

2.2 審査官引例の抽出

審査官引例は、登録特許の書誌情報として次のように記述される。審査官引例は、この書誌情報から抽出を行う。

(56)【参考文献】
【文献】特開 平3-161290(JP,A)
【文献】特開 昭61-211177(JP,A)
【文献】特公 昭48-39425(JP,B1)

2.3 予備実験における引例抽出結果

国内特許の引例には発明者引例、審査官引例とも、国内特許以外に外国特許や文献(論文、本、...)が記述される場合がある。そこで、どのような引例がどの程度存在するかを調査する予備実験として、1993～2003年の登録系特許1,235,185件を対象に上記方法で引例抽出を行った。

予備実験における抽出引例数を、表1に示す。「特許件数」は1件以上の発明者引例を持つ特許件数を、「平均引例数」は1件以上の発明者引例を持つ特許における平均引例数をそれぞれ表す。

表 1: 予備実験における抽出引例数

引例	引例数	特許件数	平均引例数
発明者引例	1,899,398	530,566	3.6
審査官引例	3,138,601	985,968	3.2

発明者引例は全体の 43% (= 530,566 / 1,235,185) の特許で 1 件以上記述されているのに対し、審査官引例は 80% (= 985,968 / 1,235,185) とかなりの率で付与されている。

発明者引例、審査官引例とも引例が記述される場合はどちらも平均 3 件ほどである。

発明者引例と審査官引例の重なりは 142,646 件と、あまり多くはなかった(発明者引例の 8%、審査官引例の 5%)。

発明者引例に関する抽出引例数の、国内特許、外国特許、文献別の内訳を表 2 に示す。「列挙型」は列挙型の記述から抽出された引例の内数である。内訳の例えば「特開」は、特許明細書において「特開昭 5 0 - 1 1 7 4 1 4」のように記述された引例を表す。発明者引例の場合、文献は抽出対象には含めていない。

表 2: 発明者引例の内訳

引例数	割合	列挙型	内訳
860,180	45.3	119,111	特開
483,627	25.5	175,825	米国特許
196,402	10.3	15,785	特公
101,170	5.3	8,106	特願
69,139	3.6	535	実開
34,922	1.8	8,357	英国特許
28,465	1.5	200	実公
11,856	0.6	1,002	欧州特許
7,965	0.4	582	特許
6,709	0.4	1,591	西独特許
5,475	0.3	28	実願
...

列挙型の記述は「特開」の 14% (= 119,111 / 860,180) のように国内特許では概ね 1 割程度であることが多いのに対し、米国特許で 36% (= 175,825 / 483,627)、英国特許で 24% (= 8,357 / 34,922) と外国特許の場合に多くみられる。

なお 1993 ~ 2003 年の公開系特許 3,817,234 件を

対象に同様な実験を行ったが、結果は大体同じ傾向であった。

審査官引例に関する抽出引例数の内訳を表 3 に示す。

表 3: 審査官引例の内訳

引例数	割合	内訳
2,395,939	76.3	特開
533,919	17.0	実開
118,267	3.8	特公
63,971	2.0	その他(文献など)
63,887	2.0	実公
37,774	1.2	米国特許
24,581	0.8	特表
11,224	0.4	国際公開
9,861	0.3	特許
5,232	0.2	欧州特許出願公開
5,131	0.2	登録実用新案
...

表 2 と表 3 を比較すると、発明者引例には外国特許が多く、審査官引例には実用新案が多いことが分かる。

発明者引例、審査官引例とも公開番号で記述されることが多いが、出願番号や特許番号などで記述されることもある。被引用数を正確なものとするため、番号はすべて出願番号に統一した。

外国特許や文献は、表記の揺れがあり引例数も国内特許に比べそれほど多くはないことから本稿では抽出しないこととした。

2.4 引例抽出結果

引例抽出では 1993 ~ 2004 年前半の特許 5,304,717 件を対象に、発明者引例と審査官引例を抽出した。

抽出引例数を表 4 に示す。

表 4: 抽出引例数

引例	引例数	特許件数	平均引例数
発明者引例	5,495,682	1,852,021	3.0
審査官引例	3,892,762	1,228,114	3.2

予備実験の審査官引例の付与率が80%だったのに比べここでは23% (= 1,228,114 / 5,304,717) と低いのは、審査官引例は登録特許のみに付与されるためと思われる。

次に、被引用数に関する結果を示す。

1回以上発明者引例により引用されている特許件数は1,765,854件、審査官引例により引用されている特許件数は2,548,534件であった。発明者引例と審査官引例の被引用数の分布を表5に示す。

表5: 被引用数の分布

被引用数	発明者引例	審査官引例
1 ~ 10	1,710,044	2,544,540
11 ~ 20	31,662	3,662
21 ~ 30	8,883	253
31 ~ 40	3,980	40
41 ~ 50	2,329	23
51 ~ 60	1,501	6
61 ~ 70	1,137	5
71 ~ 80	842	5
81 ~ 90	619	0
91 ~ 100	441	0
101 ~	4,416	0

表5より発明者引例の方が被引用数の多い特許が多く存在することから、発明者引例の方が審査官引例よりも特定の特許に被引用が集中していることが分かる。

予備実験における「発明者引例は外国特許が多く、審査官引例は実用新案が多い」という結果と、この特定の特許への集中傾向とから、発明者引例の方が審査官引例よりも対象分野のトレンドなど技術的な面を強く反映する傾向にあると推測される。

3 引例に基づく重要特許抽出

本節では、引例に基づく重要特許抽出について述べる。

文献と同様に、特許の重要度と被引用数はある程度の関連があると言われている [1]。そこで、前節で抽出した引例に基づき被引用数の大きな特許の抽出を行い、実際に重要特許を抽出することが出来るかどうか実験を行った。

3.1 発明者引例と審査官引例

重要特許抽出において発明者引例と審査官引例のどちらを用いるかについてであるが、前節で述べたように発明者引例の方が審査官引例よりも技術的な面を強く反映する傾向にあると思われるので、基本的に発明者引例を用い、適宜審査官引例と比較することにする。

3.2 自社引用と他社引用

引用分析においては自社引用と他社引用を区別し、特許の重要度を評価するような場合は他社引用を用いるということがしばしば行われる。しかし、自社引用と他社引用の境は必ずしも明確なものではない。組織は合併や分割で時間と共に変わり得る。またグループ企業間の引用を自社引用と他社引用のどちらとするか、共出願特許に関する引用の扱いをどうするかなど、単純に自社引用か他社引用かを判断し難い場合もある。

本稿では、次の3つの値をそれぞれ特許の重要度として実験を行い比較する。出願人や発明者が複数の場合は、一組織あるいは一人でも重なりがある場合に自社引用とした。被引用数は、全特許における被引用数ではなく重要特許抽出対象である特許集合における被引用数を用いた。

- (i) 自社引用および他社引用の両方を合わせた引用の被引用数
- (ii) 発明者に基づく他者引用による被引用数
- (iii) 他社引用による被引用数

3.3 引用情報を用いた特許集合の補完

実験では、特許検索システムから検索キーにより重要特許抽出対象である特許集合を求める。検索キーによる検索だけでは検索漏れを生じ得る。そこで引用情報を用いて検索結果の特許集合の補完を行う。補完の際重要なことは、

- 重要特許を補うこと
- ノイズを入れないこと

の2点である。補完の方法としては、引用情報により引用方向、あるいは被引用方向にn段階以内の特許を追加する、あるいはさらに補完により追加される特許集合に特徴的なキーワードやIPCなどを検索キーに修正を加える、などが考えられる。

本稿では重要特許を補いつつノイズを入れない方法として、元の特許集合からの被引用数が閾値¹以上の特許のみ追加することとした。

3.4 重要特許抽出実験と評価

テーマ「歩行ロボット」について実験と評価を行った。

重要特許抽出の実験手順は次の通りである。

1. 特許検索システムから検索キーにより特許集合を求める
2. 引用情報を用いて特許集合を補完する
3. 被引用数上位の特許を重要特許として抽出する

検索キー「歩行 and ロボット」で検索して得られた 1,910 件に引用情報より補完した計 1,916 件を対象に、被引用数上位の特許を抽出した。

これに対し、特許庁サイトの「特許出願技術動向調査報告(平成 13 年度) ロボット」²に挙げられている特許を重要特許として比較を行った。

引用として発明者引例を用い、特許の重要度として (i) を用いた場合の結果を次に示す。

「第 14 図 歩行技術の変遷」に挙げられている特許 3 件は表 6 に * で示した通り被引用数 2 件以上の 59 位以内に出現した。

表 6: 歩行技術の重要特許抽出結果

被引用数	順位	特許
57	1	特願平 01-324218
29	2	特願平 04-137880
28	3	特願平 04-137886
...		
14	8	* 特願平 04-164297
...		
6	17 ~ 20	* 特願平 09-213953
...		
2	36 ~ 59	* 特願平 11-311813

「第 15 図 二足歩行の安定性向上技術の展開」の流れ図に挙げられている特許 18 件の内 12 件は表 7 の通り被引用数 2 件以上の 59 位以内に出現した。

¹元の特許集合からの被引用数の最大値の 1/5。

²<http://www.jpo.go.jp/shiryu/pdf/gidou-houkoku/robot.pdf>

表 7: 二足歩行の重要特許抽出結果

被引用数	順位	特許件数
28	3	1
24	5 ~ 6	2
14	8	1
10	10	1
9	11 ~ 13	1
7	15 ~ 16	1
6	17 ~ 20	1
5	21 ~ 22	1
3	26 ~ 35	2
2	36 ~ 59	1

重要特許の多くが被引用数上位数十件以内に出現する。ただし被引用数が必ずしも十分大きい値とは言えないため、特許の重要性判定の目安のひとつとして用いるのが適当と思われる。

自社引用と他社引用に関して、特許の重要度として (ii), (iii) を用いた場合の重要特許抽出実験の結果は、いずれも (i) と同じような傾向を示した。

引用として審査官引例を用いた場合は、「第 14 図 歩行技術の変遷」に挙げられている特許 3 件はいずれも被引用数 0 であった。「第 15 図 二足歩行の安定性向上技術の展開」の流れ図に挙げられている特許 18 件については、表 8 の通り発明者引例では抽出されなかった 6 件の内 2 件が抽出された。ただしそれ以外の 16 件は被引用数 1 以下であった。このことから、発明者引例と審査官引例とを相補的に用いることがひとつの有効な手段と思われる。

表 8: 二足歩行の審査官引例による抽出結果

被引用数	順位	特許
8	1 ~ 2	* 特願平 01-297199
8	1 ~ 2	特願平 01-324217
6	3	特願平 09-209306
5	4	* 特願平 01-167295

3.5 引用情報の時系列的性質

特許などの被引用数は公開されてからの期間が長いほど大きくなる。そこで特許を出願年毎に並べ、時系列的性質を観察した。

テーマ「歩行ロボット」関連特許の出願年と被引用数のグラフを図 1 に示す。引用としては発明者引例を用いた。グラフにおいて、「平均」は平均被引用数を、「歩行技術」は「第 14 図 歩行技術の変遷」

に挙がっている特許 3 件を、「二足歩行」は「第 15 図 二足歩行の安定性向上技術の展開」の流れ図に挙がっている特許 18 件を、「上位」は被引用数 2 件以上の 59 件の特許をそれぞれ表す。また「件数」は特許件数を右側の軸で表す。

重要特許の被引用数はその多くが平均被引用数よりも大きいことが分かる。公開されてからの期間を考慮して被引用数の評価を行うことで重要特許抽出の性能向上につながる可能性があり、これは今後の課題である。

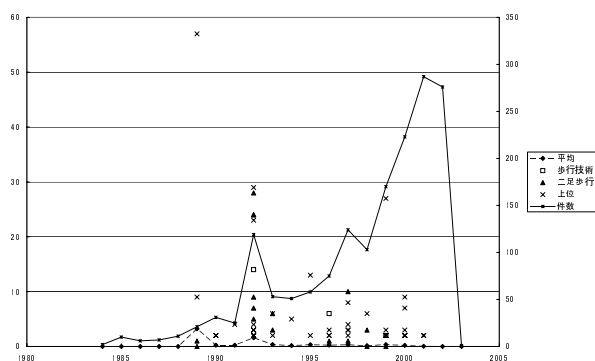


図 1: 「歩行ロボット」関連特許の被引用数

4 引用情報の可視化

本節では引用情報に基づく、重要特許間の関係の可視化について述べる。

4.1 可視化のアプローチ

引用情報の可視化の従来の研究は、共引用などに基づき文献間や著者間の静的な関係を表すものが多い [2], [3]。このアプローチでは時系列的な変化や流れが表現されにくいという欠点がある。

別のアプローチとして Hyperbolic Tree[4] のように時間の流れに沿った可視化手法がある。だが多くの手法は引用情報をそのまま可視化するだけであり、そのままでは表示が煩雑となり可視化対象となっている文書集合全体の特徴を読み取ることは難しいことが多い。

我々は先に、文書集合全体の特徴を表すような時系列的な流れを表現する可視化手法を提案した [5]。ここでは文書をノード、引用をエッジとする有向グラフである引用フローにより可視化し、エッジ絞込みおよびエッジ強調を行うことで文書集合に特徴的な技術

の流れを適切かつ簡潔に表現する。

この引用フローには、同時期に出願された特許同士など本来関連のある特許同士が繋がらないことがある、という課題があった。本稿では引用以外に共引用および書誌結合をエッジとして加えた可視化手法を提案する。共引用は同じ文書から共通して引用される文書間に関連があるとする関連性であり、書誌結合は同じ文書を共通して引用する文書間に関連があるとする関連性である (図 2)。

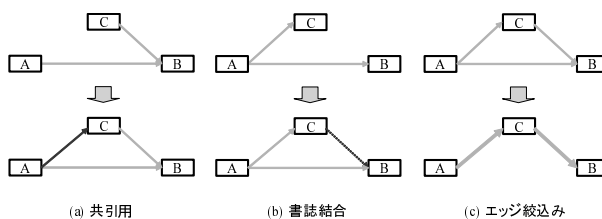


図 2: 共引用, 書誌結合, エッジ絞込み

4.2 提案手法

提案手法では、ある一定値以上の共引用スコアや書誌結合スコアの共引用や書誌結合をエッジとして加え、その後エッジ絞込みを行うことで文書フローの簡潔化を行う。

2 特許 A, B 間の共引用スコア S の定め方としては、予備実験より被引用数と比率の両方が考慮された値が比較的妥当な可視化が得られたことから、次のような相互情報量を用いる。書誌結合スコアの定め方も同様である。なお、 N_{AB} は A, B を共に引用している件数を、 C_A は A の被引用数を、 C_B は B の被引用数を、 N は全特許件数を、それぞれ表す。

$$S = P_{AB} \times \log \frac{P_{AB}}{P_A \times P_B}$$

$$P_{AB} = \frac{N_{AB}}{N}$$

$$P_A = \frac{C_A}{N}$$

$$P_B = \frac{C_B}{N}$$

4.3 引用フロー作成実験と評価

本提案手法の評価としては、前出の「特許出願技術動向調査報告 (平成 13 年度) ロボット」の「第 15 図 二足歩行の安定性向上技術の展開」の流れ図と、本提案手法による引用フローとの比較を行う。可視化対象

の特許が異なると比較が難しいため、引用フローの対象特許は流れ図に挙がっている特許 18 件を用い、引用フローのエッジのつながり方に注目した比較を行う。

二足歩行の流れ図を我々の可視化ツールを用いて表現したものを図 3 に示す。ノードを囲む線とその上の「足首制御」などは、評価のために後から書き足したものである。図 3 から読み取れる主な技術の流れには次のようなものがある。

- 足首制御と両脚制御からハイブリッド制御へと発展した
- 両脚制御と上体制御などから、上体と脚部の協調や全身協調制御へと発展した

次に、「ロボット」関連特許の全引用を可視化したものを図 4 に示す。引用としては発明者引例と審査官引例の両方を用いた。図で二重枠のノードで示した特許「特願平 04-253878」は、図 3 の位置のままでは線の交差が多く見づらいため位置を変更した。

図 4 を見るとエッジが多いため多少煩雑で、情報を読み取るのはやや難しい。さらにノード数が増えることにより煩雑になることが予想され、何らかの簡潔化操作が望まれる。本提案手法はノード間の連結性を保存したままフローを簡潔化することが出来る。

最後に本提案手法により、図 4 に共引用と書誌結合を表すエッジを追加し、エッジ絞込みを行った引用フローを図 5 に示す。図 5 において、共引用は黒い実線矢印で、書誌結合は黒い点線矢印で、それぞれ示す。

共引用により両脚制御に関する「特願平 04-137884 ~ 6」を結ぶエッジが追加される。書誌結合により足首制御に関する「特願平 04-137881」と両脚制御に関する「特願平 04-137884」を結ぶエッジが、そしてまたハイブリッド制御に関する「特願平 10-032420」、「特願平 10-125232 ~ 3」を結ぶエッジがそれぞれ追加される。エッジ絞込みにより、これら新たに追加されたエッジの周辺が簡潔化される。

図 4 に比べエッジが減ったため図がすっきりし、図 3 から読み取れたのと同じような技術の流れを読み取ることが出来る。

図 3 との違いとしては、特許「特願平 04-253878」の位置付けや、図 3 ではつながっている特許間が引用フローではつながっていない所がある、などが挙げられる。

このようにいくつかの違いはあるものの本提案手法を用いることで、人手による流れ図とほぼ同じよう

な技術の流れが読み取れるフローを自動で作成することが出来た。特定のテーマに関する重要な技術とそれらの間の関係の把握や、技術の流れ図の作成支援において、本提案手法は有効であると考えられる。

5 まとめ

国内特許約 530 万件から発明者引例、審査官引例の抽出を行い、それぞれの性質の違いを調べ、発明者引例の方が技術的な面を強く反映する傾向にあることを明らかにした。

抽出した引例のうち主に発明者引例を用いて、引例の被引用数に注目することで重要特許の多くを見付けることが出来た。

重要特許間の関係を知るための引用情報に基づく可視化手法として、共引用や書誌結合をエッジに加え、エッジ絞込みにより煩雑さを軽減させる引用フローを提案し、その有効性を確認した。

重要特許抽出の性能向上に向けた今後の課題として、整理標準化データの利用、外国特許や文献とのリンク、明細書のテキスト情報の利用などが挙げられる。

参考文献

- [1] 富澤宏之：特許解析で見る技術開発動向 引用分析を中心として，*CICSJ Bulletin*, Vol. 16, No. 6, pp. 17-20 (1998).
- [2] Small, H.: Visualizing Science by Citation Mapping, *J. Am. Soc. Information Science*, Vol. 50, No. 9, pp. 799-813 (1999).
- [3] Chen, C. and Paul, R. J.: Visualizing a Knowledge Domain's Intellectual Structure, *IEEE Computer*, Vol. 34, No. 2, pp. 65-71 (2001).
- [4] Lamping, J., Rao, R. and Pirolli, P.: A Focus + Context Technique Based on Hyperbolic Geometry for Visualizing Large Hierarchies, *ACM Conference on Human Factors in Computing Systems (CHI '95)*, pp. 401-408 (1995).
- [5] 小川知也, 渡部勇：技術文書引用情報の可視化, 情処研報 情報学基礎 70, pp. 67-74 (2003).

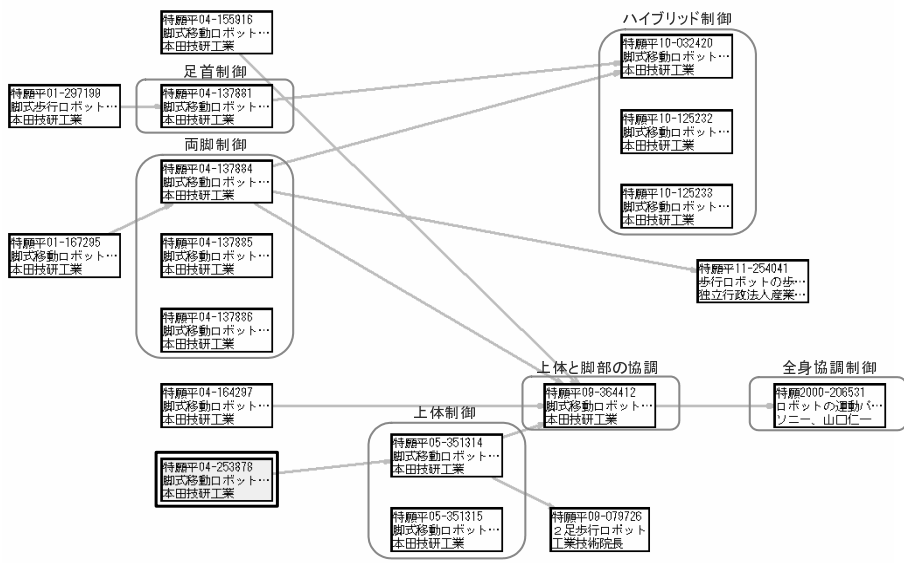


図 3: 二足歩行の流れ図

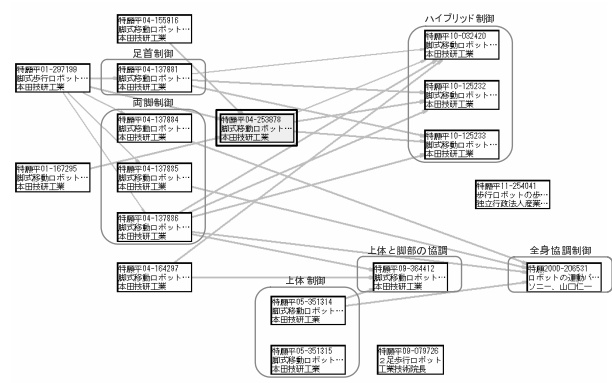


図 4: 「ロボット」関連特許の全引用

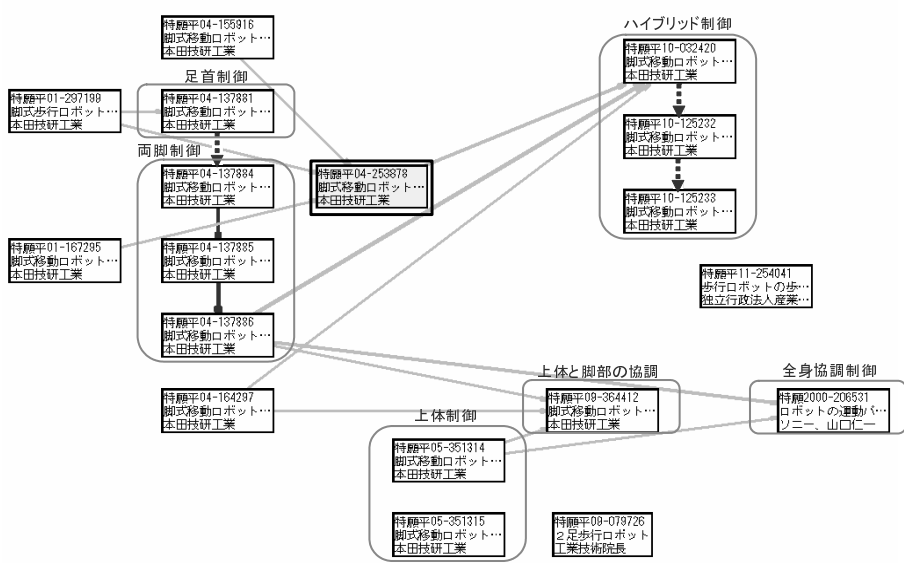


図 5: 「ロボット」関連特許の引用フロー