L1C10 - Tool 2: Search and example data

0:01	in the last concept we finished with	私たちが完成した最後のコンセプトでは
0:02	remote procedure calls where	**リモートプロシージャコール**の場所
0:04	someone who's a data scientist can can	**データサイエンティスト**ならできる
0:06	leverage a very familiar interface to	非常に使い慣れたインターフェイスを活用して、
0:09	control data science tools on remote	データサイエンスツールをリモートで**制御**
0:11	machines thus making it so they don't	機械はこうしてそうならないようにするのです
0:13	have to centralize all the data in one	すべてのデータを 1 つに集中管理する必要がある
0:14	location and avoiding some of the more	**場所を特定し**、さらに**いくつかのことを回避する**
0:16	challenging aspects of acquiring a copy	コピーを入手する際の困難な側面
0:18	of a data set but we finished this	データセットの一部ですが、これで完了しました
0:20	section with with this interesting	これは興味深いセクションです
0:22	question of how can we actually do good	実際にどうすれば良いことができるのかという問題
0:24	data science without looking at the data	データを見ないデータサイエンス
0:26	ourselves	私たち自身
0:27	in this concept we're going to talk	このコンセプトで話します
0:29	about	だいたい
0:30	the product experience of	の製品体験
0:33	of privacy and technologies meaning like	**プライバシー**と**テクノロジー**の意味
0:35	what are the things that we can do to	そのために私たちにできることは何でしょうか
0:37	make it	成功する
0:38	feel like we have a copy of the data	データのコピーを持っているような気がする
0:39	even if we don't or to try to bridge	たとえそうでなくても、橋渡しをしようとしても
0:41	those gaps where there are where the	そこにあるギャップは、
0:43	limitations and the real theme that	限界と本当のテーマ
0:44	we're going to see here is that	ここで見てみましょう
0:46	that with with tools like search or	**検索**や
0:49	tools like example data	**サンプルデータなどのツール**
0:51	and then there's other tools that we can	他にもできるツールがあります
0:52	go into as well we're going to focus on	にも焦点を当てます
0:53	these two	この二つ
0:54	there's this whole legacy of tools for	ツールのこの遺産全体が存在します
0:57	working with data that's simply too big	単純に大きすぎるデータを扱う
1:00	for you to read all of it and it turns	あなたが**それをすべて読むと**、それが変わります
1:02	out this sort of experience translates	この種の**経験が翻訳される**
1:04	really nicely into data that's simply	単純にデータにうまく取り込むことができます
1:06	too sensitive or too distant for you to	あなたには敏感すぎる、または遠すぎる
1:08	actually see all of it so so in the case	実際にすべてを見てみましょう
1:10	of of a data set that's too large for	大きすぎるデータセットのうちの
1:12	you to to to find relevant data points	関連するデータポイントを見つけるには
1:15	you might use a search tool to actually	実際に検索ツールを使用するとよいでしょう
1:17	iterate through millions and millions	何百万も繰り返す
1:18	and millions of images looking for you	そして何百万もの画像があなたを探しています
1:20	know candid examples of one one pattern	一つ一つのパターンの率直な例を知る
1:22	or another or you might look at	または別のもの、またはあなたは見るかもしれません
1:23	individual sub samples of the data as a	データの個々のサブサンプルを
1:26	sort of representative sample of this	これの代表的なサンプルのようなもの
1:28	big big big big data that you can't look	大きな大きな大きな大きなビッグデータは見ることができません
1:30	at yourself so it's really the theme	自分自身にそれが本当にテーマです
1:31	that we're going after and i hope you	私たちはそれを追いかけています、そして私はあなたがそれを願っています
1:32	sort of see that this transfers over	これが転送されるのがわかります
1:34	quite nicely i'm going to walk through a	とてもうまく、私は通りを歩くつもりです
1:36	couple example interfaces of this being	この**存在のインターフェイス**のいくつかの例
1:38	the case so let's say we've got an	ケースがあるので、
1:39	interface to remote a remote data	**リモートへのインターフェース**、リモートデータ
1:41	science portal	科学ポータル
1:43	and um you know i've got this this	そして、ええと、私はこれを持っていることを知っています
1:46	client towards this this remote pi grid	このリモート PI グリッド**に対するクライアント**
1:49	server	サーバ
1:50	and so i call grid.search right standard	それで私は**grid.search**を正しい標準と呼びます
1:54	standard idea and the thing is that i'm	標準的な考え、そして**問題は私がそうである**ということです

1:55	actually searching on the public	実際に一般公開で検索している
1:57	metadata in this case	この場合のメタデータ
1:59	to be able to find private data that's	プライベートデータを見つけることができる
2:01	relevant for me right so in this case i	私にとって適切なので、この場合は私は
2:02	might be looking for diabetes data and i	糖尿病のデータを探しているかもしれないし、
2:04	find that you know there's a few few	いくつかあることを知っていることがわかります
2:06	results that were returned to me um you	私に返された結果、ええとあなた
2:08	know if i look at one of these pointers	これらのポインタの 1 つを見ればわかる
2:09	there's sort of certain metadata that	ある種のメタデータがあります
2:11	comes with it right so i can see you	ちゃんと付属しているので、会えますよ
2:13	know some tags i can see shape i can see	いくつかのタグを知っています、形が見えます、見えます
2:15	you know some descriptions of how this	あなたはこれがどのように起こるかについていくつかの説明を知っています
2:17	data was collected what the data is	データは収集されました データは何ですか
2:18	actually about	実は約
2:20	and of course you know attach this data	もちろん、このデータを添付することはご存知でしょう
2:22	can be other public data that tells	それが伝える他の公開データである可能性があります
2:24	me even more about this information such	この情報についてさらに詳しく知りたい
2:26	as such a sample data so i can actually	このようなサンプルデータなので、実際に
2:28	look at something and sort of understand	何かを見てなんとなく理解する
2:30	some of the latent patterns that that	潜在的なパターンのいくつかは、
2:32	that might not be obvious from the	それは、からは明らかではないかもしれません
2:33	description in terms of how the data set	**データセット**の方法に関する説明
2:34	works um and so what i this is really	**うまくいきます**、それで、これは実際には何ですか
2:37	just kind of like a a small nugget of of	の小さな塊のようなもの
2:40	a broader product experience	より幅広い製品体験
2:42	uh problem but with but for which there	うーん、問題はあるけど、それはそれで
2:45	are many many solutions and i think that	解決策はたくさんありますが、私はそう思います
2:47	there's there's a longer period of time	もっと**長い期間がある**
2:50	through which the sort of remote	それを通して一種の**リモ**コンが
2:51	data science experience is going to	**データサイエンス**の経験**は**、
2:53	continue to iterate but we have lots and	**繰り返しを続けます**が、たくさんあります
2:54	lots of really good fodder to start with	まずは本当に良い飼料がたくさんある
2:56	simply because the problem of doing data	単純にデータ処理の問題があるからです
2:58	science and data that that you're not	科学とデータによると、あなたはそうではない
2:59	allowed to actually look at in its	実際に見ることができます
3:00	entirety is very similar to the product	全体が製品と非常によく似ています
3:03	experience of working with data that's	データを扱った経験
3:05	just too big for you to ever really	あなたには本当に大きすぎる
3:06	think about looking at in its entirety	全体を見ることを考える
3:08	right	右
3:09	so what tools have we looked at so far	これまでにどのようなツールを見てきましたか
3:11	remote procedure calls where data	**リモート プロシージャ コール**のデータ
3:12	remains in our remote machine that	リモートマシンに残っているのは、
3:13	solves some really big important	本当に重要ないくつかを解決します
3:15	problems right that does most of the	ほとんどの問題を解決します
3:16	heavy lifting of sort of remote data	リモートデータのような重労働
3:18	science	化学
3:19	second is search and sample data you	2 番目は**検索**と**サンプル データ**です。
3:20	know we can we can feature engineer with	フィーチャ エンジニアができることを知っています
3:22	sample data you know and if we actually	あなたが知っているサンプルデータと、実際に
3:24	need to look at some pieces we could we	いくつかの作品を見る必要があるので、できますか
3:25	could request small snippets from from	から小さなスニペットをリクエストできます
3:27	the data owner to actually let us look	データ所有者に実際に見てもらいましょう
3:29	at it or maybe like a synthetic	あるいは合成のようなものかもしれません
3:31	generator like a gan or something like	ガンなどのジェネレーター
3:32	this can can generate similar data but	これにより同様のデータを生成できますが、
3:34	not quite exactly the same data that is	まったく同じデータではありません
3:37	that is the actual private data at a at	それは、**ある時点の実際のプライベートデータ**です
3:39	a remote location um but we still	遠い場所、ええと、でも私たちはまだ
3:41	haven't solved all the problems so for	すべての問題を解決したわけではないので、
3:43	example	例
3:44	we can still	まだできます
3:45	**pull data out using this dot get method**	**この dot get メソッドを使用してデータを取り出します**
3:48	right what does it actually mean for us	それは私たちにとって実際に何を意味するのでしょうか

3:50	to to	へ へ
3:52	ask for our results back in a secure way	**安全な方法で結果を要求する**
3:55	like this seems like a really obvious	これは本当に明らかなことのように思えます
3:57	bottleneck and if i've got pointers to	ボトルネックとそのヒントがあるかどうか
3:59	the remote data can i just call dot get	**リモート**データは **dot get を呼び出す**だけでよいでしょうか
4:01	on those pointers and actually pull the	これらの**ポインターに基づいて実際にプル**します
4:03	data out myself	自分でデータを出します
4:04	the next concept we're going to explore	私たちが検討する次のコンセプト
4:06	a really powerful technology called	と呼ばれる非常に強力な**テクノロジー**
4:08	differential privacy which can be used	利用できる**差分プライバシー**
4:09	to address this problem see you then	この問題を解決するには、また会いましょう

英語 (自動生成)

OpenMined https://www.youtube.com/watch?v=D9j_WCT25Zg