

BAB 2

LANDASAN TEORI

2.1 Ambigu

Berdasarkan Kamus Besar Bahasa Indonesia, ambigu berarti mempunyai makna lebih dari satu. Keambiguan ini dapat menimbulkan keraguan atau ketidakjelasan dalam kalimat yang diucapkan atau ditulis [4].

2.1.1 Jenis Jenis Ambigu

Ambigu terbagi menjadi beberapa jenis berdasarkan bentuknya, yaitu ambiguitas fonetik, ambiguitas gramatikal, dan ambiguitas leksikal.

1. Ambiguitas Fonetik

Jenis ambigu yang pertama yaitu ambiguitas fonetik. Jenis keambiguan ini terjadi akibat persamaan bunyi yang diucapkan. Karena keambiguan ini terjadi saat percakapan, maka jenis ini sering terjadi dalam dialog sehari-hari.

Contoh:

Putri datang ke sini memberi tahu.

Frasa “memberi tahu” dapat mengandung dua arti yaitu memberi tahu (makanan yang terbangun dari kedelai), atukah memberikan suatu informasi. Keambiguan ini muncul karena bunyi yang diucapkan antara “memberi tahu” yang berarti memberikan makanan dan “tahu” yang berarti “memberikan informasi” sama. Oleh karena itu perlu didengarkan pembicaraan secara lengkap.

2. Ambiguitas Gramatikal

Sesuai dengan namanya, ambiguitas gramatikal terjadi karena proses pembentukan ketatabahasaannya. Akan tetapi, kata-kata yang mengalami ambiguitas jenis ini akan hilang jika sudah masuk dalam konteks kalimat.

Contohnya penggunaan kata orang tua. Kata tersebut memiliki dua arti yaitu, orang yang sudah tua, dan ibu bapak. Ketidakjelasan ini akan sirna setelah adanya kalimat berikut.

- a. Orang tua kandung Budi belum diketahui keberadaannya. (makna: ibu bapak)
- b. Kemarin maghrib aku bertemu orang tua bertongkat hitam. (makna: orang yang sudah tua)

3. Ambiguitas Leksikal

Jenis ambigu yang ketiga adalah ambiguitas leksikal. Keambiguan jenis ini disebabkan oleh faktor kata itu sendiri.

Contoh:

- a. Anton berlari dengan sangat kencang ketika lomba maraton.
- b. Anton lari dari kenyataan hidup yang pahit.

Kata “lari” pada kedua kalimat di atas memiliki beda makna. Pada kalimat pertama, “lari” berarti aktivitas lari, sedangkan kalimat kedua “lari” berarti menjauh.

2.2 Natural Language Processing

Pemrosesan bahasa alami atau yang biasa disebut Natural Language Processing (NLP), merupakan sebuah titik temu dari ilmu bahasa, ilmu komputerisasi, teknik informatika, dan kecerdasan bangunan [5].

2.2.1 Word Sense Disambiguation

Word Sense Disambiguation adalah suatu proses mengidentifikasi makna kata yang digunakan dalam suatu konteks ketika kata memiliki sejumlah makna yang berbeda. [6]. Pada penelitian ini permasalahan *word sense disambiguation* menjadi titik utama permasalahan, dimana penulis sendiri sebagai peneliti merancang dan membangun sebangun aplikasi anotasi yang akan memudahkan pembangunan *annotated corpus word sense disambiguation*, dari

data masukan yang berupa dokumen teks artikel berita yang berasal dari portal berita bahasa Indonesia.

2.2.2 *Corpus*

Dalam bidang linguistik, *corpus* (jamak : *corpora*) atau *text corpus* adalah kumpulan teks dalam jumlah yang besar dan terstruktur (biasanya disimpan dan diproses secara digital). *Corpus* seringkali digunakan untuk melakukan analisis statistik dan pengujian hipotesis, memeriksa kejadian atau memvalidasi aturan linguistik dalam wilayah bahasa tertentu [7]. Pada penelitian ini, corpus yang diambil sebagai data masukan merupakan hasil dari proses *web-crawling* dengan sumber portal berita bahasa Indonesia.

2.2.3 *Annotation*

Secara umum anotasi (*Annotation*), adalah sebuah proses menambahkan komentar, atau penjelasan atas suatu dokumen teks. Khususnya pada bidang NLP, anotasi memegang peran penting sebagai cara menyimpan dan mengasosisasikan suatu label (metadata) tertentu, dari dokumen teks yang diproses [8].

Dalam dunia komputerisasi, khususnya bidang NLP, anotasi sangatlah berperan penting, contohnya saja dalam penyelesaian masalah seperti, keefektifan *search engine*, konteks bahasa natural yang tidak dimengerti oleh perangkat lunak komputer, bagaimana algoritma NLP mempelajari sentimen pengguna melalui teks, dan berbagai masalah lainnya [8].

Selain itu ada juga beberapa aspek dari anotasi yang dikenal dalam bidang NLP, antara lain :

1. Struktur bahasa (kelas kata, *syntax*).
2. Makna kata (*word sense disambiguation*).
3. Interpretasi makna kata (*semantics*).
4. Struktur dokumen (*discourse*).
5. *Speech sound*, dan *part of words/speech* (*phonetics, morphology*).

Tentu saja aspek dari anotasi yang digunakan pada penelitian ini adalah aspek makna kata (*sense/meaning of words*), dimana aplikasi anotasi digunakan

untuk melakukan pelabelan pada kata ambigu serta memberi makna yang sebenarnya pada kata tersebut sesuai dengan konteks kata tersebut dari dokumen data masukan.

2.2.4 Verifikasi dan Validasi

Verifikasi dan validasi dari sistem perangkat lunak, memiliki banyak sekali defensi berbeda. Namun defenisi yang paling umum adalah :

1. Verifikasi : memastikan perangkat lunak secara bener mengimplementasikan fungsi tertentu, sehingga memenuhi spesifikasi perangkat lunak tersebut.
2. Validasi : menentukan bahwa sistem memenuhi kebutuhan pengguna.

Defenisi tersebut kemudian diperiksa ulang dalam rangka membedakan antara perangkat lunak konvensional, dan *intelligent systems*. Sebuah *intelligent system* dibangun berdasarkan interpretasi dari daerah permasalahan, dengan ekspektasi bahwa sistem akan bertindak sesuai dengan tingkah laku seorang ahli dalam daerah permasalahan yang tersebut. Bagaimana sebuah sistem bertindak layaknya seorang manusia, menjadi sebuah tolak ukur untuk menilai sebuah *intelligent system* [9].

Defenisi sebelumnya dari, verifikasi dan validasi, dapat diaplikasikan kepada *intelligent systems*. Namun sangat dibutuhkan sedikit modifikasi dari defensi tersebut mengingat adanya *knowledge base* dan perlunya pembandingan performa sistem dengan performa manusia terhadap area permasalahan yang dibahas [9]. Permasalahan utama pada validasi dan verifikasi dari sebuah *intelligent systems* berasal dari satu tujuan yang sederhana: memastikan bahwa hasil sebuah sistem dapat menyediakan jawaban, solusi atau bertindak sesuai dengan bagaimana seorang ahli bertindak terhadap data masukan.

Maka pada tahun 2000 Barr, dan Gonzalez mendefinisikan ulang verifikasi dan validasi sehingga sesuai dengan aspek dari intelligent sistem :

1. Verifikasi :
 - a. Proses memastikan bahwa intelligent system yang dimaksud susesuai dengan spesifikasinya
 - b. Memastikan *knowledge base* konsisten.
2. Validasi : Proses memastikan bahwa data keluaran dari sebuah intelligent sistem setara dengan, data keluaran seorang ahli di bidangnya jika diberi data masukan yang sama.

Pada penelitian ini validasi diimplementasikan sebagai aspek fungsi utama dari sistem yang dibangun, dimana sistem yang dibangun akan bertujuan mempermudah seorang ahli bahasa dalam melakukan anotasi terhadap data masukan (teks artikel bahasa Indonesia), yang kemudian *output* dari sistem dapat digunakan sebagai data pembanding dalam proses validasi oleh *intelligent system* yang akan atau telah dibangun pada penelitian lainnya.

2.3 Semiotika

Secara etimologi, istilah semiotik dan semiotika berasal dari bahasa Yunani semeion yang berarti tanda. Tanda itu sendiri didefinisikan sebagai suatu yang atas dasar konvensi sosial yang terbangun sebelumnya dapat dianggap mewakili sesuatu yang lain. Tanda pada awalnya dimaknai sebagai suatu hal yang menunjuk pada adanya hal lain [10]. Pada penulisan ini semiotika digunakan untuk menjelaskan visualisasi data, khususnya penandaan kata ambigu yang dilakukan pada proses anotasi kata ambigu.

2.3.1 Studi Mengenai Tanda Menurut Peirce

Upaya klasifikasi yang dilakukan oleh Pierce terhadap tanda memiliki kekhasan meski tidak bisa terbilang sederhana. Pierce membedakan tipe-tipe tanda meliputi ikon, indeks, dan simbol yang didasarkan atas relasi diantara representasi objek dan objeknya. Tipe tipe penandaan dapat dibahas secara lebih rinci sebagai berikut.

1. Ikon adalah tanda yang mengandung kemiripan “rupa” sehingga tanda itu mudah dikenali oleh para pemakainya. Didalam ikon hubungan antara representamen dan objeknya terwujud sebagai kesamaan dalam beberapa kualitas. Contohnya sebagian rambu lalu lintas merupakan tanda ikonik karena menggambarkan bentuk yang memiliki kesamaan dengan objek yang sebenarnya.
2. Indeks adalah tanda yang memiliki keterkaitan fenomenal atau eksistensial diantara representamen dan objeknya. Didalam indeks hubungan antara tanda dan objeknya bersifat kongkrit, actual, dan biasanya melalui suatu cara yang sekuensial atau kasual. Contoh jejak kaki diatas permukaan tanah, misalnya merupakan indeks dari seseorang atau binatang yang telah lewat disana, ketukan pintu merupakan indeks dari kehadiran seseorang “tamu” dirumah kita.
3. Simbol merupakan jenis tanda yang bersifat arbiter dan konvensional sesuai kesepakatan atau konvensi sejumlah orang atau masyarakat [10].

2.4 Visualisasi Data

Data yang sangat banyak, kompleks, dan tidak teratur sangat sulit untuk dimengerti sekilas pandang oleh manusia. Membaca data akan menjadi masalah ketika data secara inkremental bertambah atau berubah. Tujuan visualisasi data adalah menunjukkan mana hal-hal yang penting secara jelas, melihat pola, dan menunjukkan informasi dari berbagai dimensi. Visualisasi yang efektif membantu pengguna dalam melakukan analisis, mempermudah menyimpulkan data, dan membangun data yang kompleks menjadi lebih mudah untuk dimengerti .

Visualisasi data tidak semata-merta mengubah data menjadi grafik visual, visualisasi data memerlukan perencanaan. Setiap jenis data memiliki teknik visualisasi berdasarkan kebutuhannya. Berdasarkan tingkat kompleksitas data, untuk menghasilkan solusi yang berharga diperlukan dari berbagai disiplin ilmu, yaitu: statistika, data mining, desain grafis, dan visualisasi informasi.

Dapat disimpulkan bahwa visualisasi data merupakan cara efektif menyajikan data yang detail menjadi informasi yang mudah diterima adalah dengan cara abstraksi menjadi informasi visual [11].

Pada penelitian ini visualisasi data digunakan sebagai analisis, dan perancangan bagaimana sistem dapat menyajikan data anotasi menjadi sebuah informasi visual yang mudah dimengerti pengguna sehingga dapat membantu pengguna dalam proses anotasi, dan validasi.

2.5 Kamus Sinonim

Kamus sinonim atau sinonimkata.com adalah kamus tesaurus bahasa Inggris dan bahasa Indonesia. Yang dapat diakses melalui laman web <https://www.sinonimkata.com/>, sinonimkata.com sendiri menyediakan fungsi pencarian sinonim yang pada penelitian ini digunakan sebagai alternatif pencari sinonim kata yang pada fungsi rekomendasi makna pada anotasi kata ambigu.

2.6 Kateglo

Kateglo adalah aplikasi dan layanan web sumber dan isi terbuka untuk kamus, tesaurus, dan glosarium bahasa Indonesia. Namanya diambil dari akronim unsur layanannya: ka(mus), te(saurus), dan glo(sarium). Data dari Pusat Bahasa Departemen Pendidikan Nasional Indonesia (Pusba) merupakan hak cipta dari Pusba dan dipergunakan di Kateglo dengan seizin Pusba. Cara pengaksesan API Kateglo dapat dilihat pada Tabel 2.1

Tabel 2.1 Keterangan API Kateglo

URL Akses	http://kateglo.com/api.php?format=[xml json]&phrase=[lema_yang_dicari] .
Parameter	format=[xml json]; phrase=[kata_yang_dicari]
Return Value	JSON atau XML

Pada penelitian ini API Kateglo digunakan sebagai kamus sinonim, dimana kamus sinonim itu sendiri berperan penting dalam memberikan rekomendasi makna untuk menghilangkan keambiguitas (*dismabiguation*).

2.7 *User interface*

Sebuah sistem komputer memiliki tiga aspek yaitu perangkat keras (hardware), perangkat lunak (software), dan manusia (brainware) yang saling terkait dan berhubungan. *User interface* atau antarmuka pengguna menggunakan bentuk tampilan grafis yang berhubungan langsung dengan pengguna. Interface adalah sebuah tempat dimana interaksi antara pengguna dan sistem pada antarmuka pengguna adalah pengoperasian dan kendali sistem operasi efektif dan umpan balik dari sistem operasi yang membantu operator membangun keputusan operasional.

User interface adalah bagian dari komputer dan perangkat lunak dapat berinteraksi (melihat, mendengar, menyentuh, berbicara) dan dimengerti oleh pengguna. UI pada intinya memiliki dua komponen yaitu *input* dan *output*. *Input* adalah cara bagaimana seseorang menyampaikan kebutuhannya atau keinginannya kepada komputer, dengan perangkat *input* yang umumnya adalah keyboard dan mouse. *Output* adalah cara bagaimana komputer menyatakan hasil dari perhitungan dan kebutuhan pengguna [12]. Pada penelitian ini UI digunakan sebagai indikator terhadap analisis UI/UX, dan perancangan UI dari aplikasi *anotasi word sense disambiguation*.

2.8 *User experience*

User experience (UX) merupakan cara bagaimana layanan atau produk bekerja ketika seseorang datang atau menggunakannya, bukan mengenai kinerja internalnya. Untuk itu UX merupakan bagian yang penting dalam berbisnis. UX memiliki lima elemen dasar (Garret, 2011) antara lain [13].

1. *Strategy plane*, merupakan lapisan paling bawah yang berupa strategi bisnis mendasar yang menentukan tujuan dari produk dan kebutuhan penggunaannya.
2. *Scope plane*, merupakan translasi dari strategi untuk diteruskan dalam pembangunan spesifikasi fungsional.
3. *Structure plane*, dibentuk arsitektur informasi yang disusun untuk memfasilitasi pemahaman pengguna.
4. *Skeleton plane*, terdiri atas tiga bagian, meliputi:

- a. *Information design*, merupakan presentasi dari informasi untuk memudahkan pemahaman.
 - b. *Interface design*, merupakan pengaturan elemen antarmuka untuk memungkinkan pengguna dalam berinteraksi.
 - c. *Navigation design*, merupakan kumpulan elemen yang memungkinkan pengguna untuk menjelajahi arsitektur informasi.
5. *Surface plane*, merupakan pengalaman sensorik ke pengguna meliputi (warna, gambar, ikon) yang dibangun untuk menyelesaikan produk.

Pada penelitian ini UX digunakan sebagai indikator terhadap analisis UI/UX pembangunan aplikasi anotasi *word sense disambiguation*.

2.9 Web-crawling

Web-crawling adalah program atau rangkaian dari beberapa program yang mampu secara iterative dan otomatis mengunduh halaman web, mengekstral url, melakukan pengambilan data tertentu [14]. Sebuah aplikasi *Web-crawling* juga mampu melakukan kalkulasi sampling dalam proses *crawling*, agar dapat mengidentifikasi halaman yang diindikasikan relevan dengan dokumen yang mau diambil oleh aplikasi *web-crawling* tersebut, maupun meniadakan halaman yang sebelumnya telah di-*crawl*.

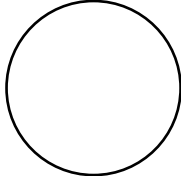
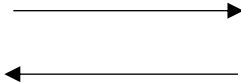
Pada penelitian ini *web-crawling* dilakukan sebagai metode pengumpulan data yang digunakan data-masukan yang nantinya diproses oleh program sebagai pembentuk corpus dan data-set word sense disambiguation. Halaman web yang merupakan target crawl dari penelitian ini adalah portal berita dalam bahasa Indonesia, dengan indeks halaman yang jelas.

2.10 Data Flow Diagram

Rosa dan Shalahuddin [15] menyatakan bahwa “Data Flow Diagram atau dalam bahasa Indonesia menjadi diagram alir data adalah representatif grafik yang menggambarkan aliran informasi dan transformasi informasi yang diaplikasikan sebagai data yang mengalir dari masukan (*input*) dan keluaran (*output*).”

Sujatmiko [16] (2012:76) berpendapat “Data Flow Diagram adalah suatu diagram yang menggunakan notas-notasi untuk menggambarkan arus dari data sistem, yang penggunaannya sangat membantu untuk memahami sistem secara logika, terstruktur, dan jelas”. Edward Yourdon dan Tom DeMarco membagi simbol atau notasi pada Data Flow Diagram menjadi 4, yang dapat dilihat pada Tabel 2.2.

Tabel 2.2 Data Flow Diagram

No.	Nama dan Simbol	Keterangan
1.	Proses 	Digunakan untuk mengfambarkan proses pengolahan data atau transfromasi data.
2.	Basis Data 	Diimplementasikan dengan pemrograman terstruktur sehingga dibuat menjadi tabel basis data yang dibutuhkan.
3.	Entitas Luar 	Digunakan untuk menggambarkan asal dan tujuan data.
4.	Alur Data 	Digunakan untuk menggambarkan aliran data yang sedang berjalan.

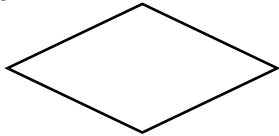
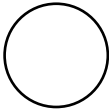
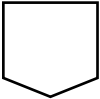
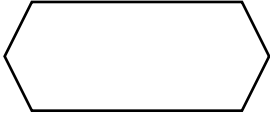
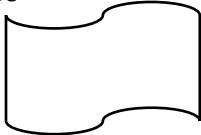
2.11 Bagan Alir (*Flowchart*)

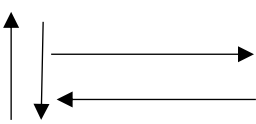
Menurut Prof.Dr. Jogyanto HM,MBA,Akt.dalam bukunya yang berjudul Analisis & Desain [17], menjelaskan bahwa, “Bagan alir (*flowchart*) adalah bagan (*chart*) yang menunjukkan alir (*flow*) didalam program atau prosedur sistem secara logika”. Bagan alir digunakan terutama untuk alat bantu komunikasi dan untuk dokumentasi.

Bagan alir sistem (*system flowchart*) merupakan bagan yang menunjukan arus pekerjaan secara keseluruhan dari sistem, bagan ini menjelaskan urutan – urutan dari prosedur – prosedur yang ada didalam sistem. Bagan alir sistem Digambar dengan menggunakan simbol – simbol yang dapat dilihat pada Tabel 2.3.

Tabel 2.3 *Flowchart*

No.	Nama dan Simbol	Keterangan
1.	Terminal 	Menyatakan permulaan atau akhir suatu program.
2.	Input / Output 	Menyatakan proses <i>input</i> atau <i>output</i> tanpa tergantung jenis peralatannya.
3	Process 	Menyatakan suatu tindakan (Proses) yang dilakukan oleh komputer.

4.	Decision 	Menunjukkan suatu kondisi tertentu yang akan menghasilkan 2 kemungkinan jawaban (ya / tidak).
5.	Connector 	Menyatakan sambungan dari proses ke proses lainnya dalam halaman yang sama.
6.	Offline Connector 	Menyatakan sambungan dari proses ke proses lainnya dalam halaman yang berbeda.
7.	Predifined Process 	Menyatakan penyediaan tempat penyediaan suatu pengolahan untuk memberi harga awal.
8.	Punch Card 	Menyatakan <i>input</i> / ouput dari / ke kartu.
9.	Punch Tape 	
10.	Document 	Mencetak keluaran dalam bentuk dokumen.

11.	Flow 	Menyatakan jalannya arus suatu proses.

2.12 Skala Perhitungan Likert

Skala pengukuran merupakan kesepakatan yang digunakan sebagai acuan untuk menentukan panjang pendeknya interval yang ada dalam alat ukur. Skala likert digunakan untuk mengukur sikap, pendapat, dan persepsi seseorang tentang fenomena sosial. Fenomena sosial ini telah ditetapkan secara spesifik oleh peneliti, dengan skala likert, maka variabel yang akan diukur dijabarkan menjadi indikator variabel. Indikator tersebut dijadikan sebagai titik tolak untuk menyusun item-item instrumen yang dapat berupa pernyataan atau pertanyaan [18].

Jawaban setiap instrumen yang menggunakan skala likert memiliki gradasi dari positif sampai sangat negatif, yang dapat berupa kata-kata sebagai berikut:

1. Sangat Setuju
2. Setuju
3. Ragu-Ragu
4. Tidak Setuju
5. Sangat Tidak Setuju

Untuk keperluan analisis kuantitatif, maka jawaban ini dapat diberi bobot, yaitu:

1. Sangat Setuju diberi 5
2. Setuju diberi 4
3. Ragu-Ragu diberi 3
4. Tidak Setuju diberi 2
5. Sangat Tidak Setuju diberi 1

Instrumen penelitian yang menggunakan skala likert dapat dibuat dalam bentuk *checklist* ataupun pilihan ganda. Contoh bentuk *checklist* dapat dilihat pada Tabel 2.4.

Tabel 2.4 Contoh Kuesioner

Berilah jawaban pertanyaan berikut sesuai dengan pendapat anda, dengan cara memberi tanda (√) atau silan (x) pada kolom yang tersedia.						
No	Pertanyaan	Jawaban				
		STS	TS	R	S	SS
1.	Apakah aplikasi yang dibangun membantu anda dalam melakukan anotasi word sense disambiguation.					
2.						

STS = Sangat Setuju diberi 1

TS = Setuju diberi 2

R = Ragu-Ragu diberi 3

S = Tidak Setuju diberi 4

SS = Sangat Tidak Setuju diberi 5

Kemudian dengan teknik pengumpulan data kuesioner, misalnya instrumen tersebut diberikan kepada 100 responden yang diambil secara random, dari 100 responden setelah dilakukan analisis di dapatkan:

25 orang menjawab SS

40 orang menjawab S

5 orang menjawab R

20 orang menjawab TS

10 orang menjawab STS

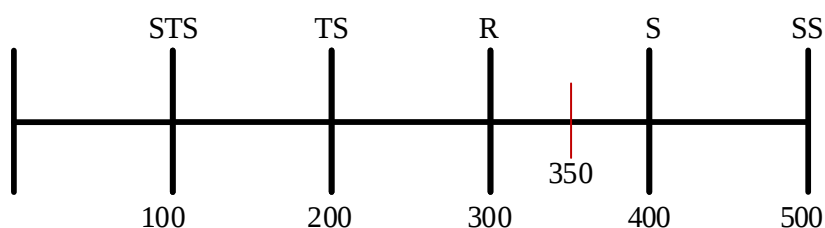
Berdasarkan data tersebut 65 orang (40+25) atau 65% responden menjawab setuju dan sangat setuju. Jadi kesimpulannya mayoritas responden setuju bahwa aplikasi

yang dibangun membantu anda dalam melakukan anotasi word sense disambiguation.

Data interval tersebut juga dapat dianalisis dengan menghitung rata-rata jawaban berdasarkan bobot setiap jawab dari responden. Berdasarkan bobot yang telah ditetapkan dapat dihitung sebagai berikut.

Jumlah bobot untuk 25 orang yang menjawab SS =	25 x 5 =	125
Jumlah bobot untuk 40 orang yang menjawab S =	40 x 4 =	160
Jumlah bobot untuk 5 orang yang menjawab R =	5 x 3 =	15
Jumlah bobot untuk 20 orang yang menjawab TS =	20 x 2 =	40
Jumlah bobot untuk 10 orang yang menjawab STS =	10 x 1 =	10
Jumlah Total =		350

Jumlah bobot ideal (kriterium) untuk seluruh item = $5 \times 100 = 500$ (seandainya semua menjawab SS). Jumlah bobot yang diperoleh dari penelitian = 350. Jadi berdasarkan data itu maka tingkat persetujuan terhadap metode kerja baru itu = $(350:500) \times 100\% = 70\%$ dari yang diharapkan (100%). Secara kontinum dapat dilihat pada Gambar 2.1.



Gambar 2.1 Contoh Skala Kontinum Perhitungan Likert

